General-Order Single-Reference and Multi-Reference Methods in Quantum Chemistry

A Thesis
Presented to
The Academic Faculty

by

Micah L. Abrams

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

School of Chemistry and Biochemistry
Georgia Institute of Technology
February 2005

General-Order Single-Reference and Multi-Reference Methods in Quantum Chemistry

Approved by:

Dr. C. David Sherrill, Advisor
School of Chemistry & Biochemistry
Georgia Institute of Technology

Dr. Thomas Orlando
School of Chemistry & Biochemistry
Georgia Institute of Technology

Dr. Mostafa A. El-Sayed
School of Chemistry & Biochemistry
Georgia Institute of Technology

Dr. Turgay Uzer
School of Physics
Georgia Institute of Technology

Dr. Rigoberto Hernandez
School of Chemistry & Biochemistry
Georgia Institute of Technology

Date Approved: 2-25-2005

ACKNOWLEDGEMENTS

I would like to thank several of the theory students who made the office a enjoyable place to work: Mutasem Sinnokrot, John Sears, Berhane Temelso, Tony Tauer, Ashley Ringer, Jeremy Moix, and James Doto. Coffee time will never be the same. I wish to thank my good friends Josh and Teresa Haynes, Mark Denton, Sarah Holloway, Michael Schuurman, and Betsy Burtner for all the emotional support and entertainment they have provided over the years. I am grateful for the wonderful people at the University of Central Arkansas Chemistry Department and Honors College. They supported me when I needed it the most, they challenged me to be a better person and a better scholar, and they taught me to truly search, and to re-search. I would like to thank the Martindale's and the Daigle's for taking me into their homes and treating me as one of their own. I would like to thank my adviser, David Sherrill, for allowing me to pursue my interests, when they were not necessarily in his best interest. The freedom I was given allowed me to build confidence in my own ideas and abilities. I wish to thank Edward Valeev for being a true friend and mentor since I was a summer student at the University of Georgia. He taught me to appreciate good beer (Guinness), good music (Tool), and good code (MPQC). Go Dawgs! Lastly, I am forever thankful to Mary Oates. When I was in high school, she took me aside and told me that I did not have to be a product of my situation. She convinced me that education would be the key to my success. I dedicate this thesis to her, the greatest mother, teacher, and friend.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Single-reference many-body perturbation theory and coupled-cluster theory, combined with carefully constructed basis sets, can be used to accurately compute the properties of small molecules. We applied a series of such methods and basis sets aimed at reaching the ab initio limit to determine the barrier to planarity for ethylene cation. For potential energy surfaces corresponding to bond dissociation, a single Slater determinant is no longer an appropriate reference, and the single-reference hierarchy breaks down. We computed full configuration interaction benchmark data for calibrating new and existing quantum chemical methods for the accurate description of potential energy surfaces. We used the data to calibrate single-reference configuration interaction, many-body perturbation theory, and coupled-cluster theory and multi-reference configuration interaction and many-body perturbation theory, using various types of molecular orbitals, for breaking single and multiple bonds on ground-state and excited-state surfaces. We also developed a determinant-based method which generalizes the formulation of many-body wave functions and energy expectation values. We used the method to calibrate single-reference and multi-reference configuration interaction and coupled-cluster theories, using different types of molecular orbitals, for the symmetric dissociation of water. We extended the determinant-based method to work with general configuration lists, enabling us to study, for the first time, arbitrarily truncated coupled-cluster wave functions. We used this new capability to study the importance of configurations in configuration interaction and coupled-cluster wave functions at different regions of a potential energy surface.

CHAPTER 1

INTRODUCTION

Computational quantum chemists strive to develop methods for the computation of molecular wave functions and properties derived therefrom to assist with the rationalization of experimental observations, and when possible, to make predictions. Significant advances in accuracy and system size have been made in the past twenty years, partly because of better methods for describing electron correlation and more efficient and parallel software, and because of advances in computer technology.

Quantum chemical methods can be divided into two categories, wave function-based methods and density functional theory. The 1998 Nobel Prize in Chemistry was awarded for pioneering contributions in both areas. The work presented in this thesis utilizes wave function-based methods because they can be applied in a systematic manner to approach the 'exact' solution to a given chemical problem. The present state of density functional theory does not allow strict control of the accuracy, i.e. it is not possible to systematically improve the density functional toward the exact solution.

Two recent studies [1, 2] describe state-of-the-art calculations of molecular properties. They report atomization energies converged to less than 1 kcal/mol, heats of formation to 1 kJ/mol, 0.1 pm for bond distances, and 1 cm^{-1} for vibrational frequencies. This level of accuracy requires the systematic application of coupled-cluster theory and basis sets constructed to converge to the complete basis set limit. In Chapter 2, we applied a similar strategy to obtain the ab initio limit to the vibrationless barrier to planarity of ethylene cation.

The procedures detailed in Chapter 2 assume a single configuration is an appropriate reference on which to build the hierarchy of coupled-cluster methods. However, there are many situations where the single configurational description breaks down. The primary focus of this thesis is the failure of the single-reference hierarchy for potential energy surfaces

corresponding to bond dissociation. Even in the simple case of H_2, two configurations strongly interact at stretched geometries and become degenerate at dissociation. Of course, molecules that are multiply bonded will have many more strongly-interacting electronic configurations. A quantum chemical method should be able to efficiently and, with equal accuracy, describe the electron correlation at all regions of a potential energy surface.

Multi-reference methods are currently the method of choice for accurately computing potential energy surfaces of small molecules [3]. They utilize a multi-configurational reference to account for the quasi-degenerate or strongly-interacting configurations (referred to as non-dynamical correlation), followed by configuration interaction, perturbation theory, or coupled-cluster theory to accurately treat the short-range electron repulsion (referred to as dynamical correlation). Including all single and double excitations from two reference configurations, where for example, one configuration has two electrons in the bonding orbital and the other has two in the anti-bonding orbital, means the final expansion of the wave function will include all single and double excitations in addition to a limited set of triple and quadruple excitations. Consequently, multi-reference methods are more computationally expensive than the single-reference counterpart. The computational scaling with respect to the number of configurations in the reference can be significantly decreased for multi-reference configuration interaction and perturbation theory [4–6]; however, configuration interaction does not scale properly with system size (size-extensivity error) and perturbation theory does not adequately describe the dynamical correlation required to achieve high accuracy.

Developing new methods to accurately compute potential energy surfaces is a thriving area of research [7]. Since the methods are often computationally expensive and may not translate directly to experimental measurement, comparisons are made to full configuration interaction (FCI). FCI includes all possible excitations of N electrons in M molecular orbitals and is the solution to the electronic Schrödinger equation in a given basis set. The number of configurations grows exponentially with the number of electrons or molecular orbitals restricting the application to small molecules with a modest basis set.

Chapters 3, 4, and 5 contain FCI benchmark data for spectroscopic constants and potential energy curves for several small molecules. Using the FCI data we calibrated single-reference configuration interaction, perturbation theory, and coupled-cluster theory and multi-reference configuration interaction and perturbation theory for breaking single and multiple bonds on ground-state and excited-state surfaces.

The molecular orbitals are also affected by the inadequacies of a single configurational reference. If the molecular orbitals of H_2 are determined from Hartree-Fock theory, the atomic hydrogen orbitals will not be obtained at dissociation, instead the orbitals would be intermediate between the orbitals for a hydrogen atom and those of a hydrogen anion. Usually, the molecular orbitals and the multi-configurational reference are simultaneously optimized, but the procedure is computationally expensive and often difficult to converge. In Chapter 6 we explore an alternative set of molecular orbitals for accurately computing potential energy surfaces.

In Chapter 2 we were unable to apply coupled-cluster theory with connected quadruple excitations (CCSDTQ) and in Chapters 4 and 5 we were unable to calibrate single-reference coupled-cluster theory beyond connected triples or multi-reference coupled-cluster theory of any kind because the software was unavailable (even for purchase). Beyond CCSD, the coupled-cluster equations become tedious to derive and time consuming to program and arbitrarily high levels of excitation are required to develop a multi-reference coupled-cluster code for general reference spaces. Chapter 7 describes our implementation of a general-order single-reference and multi-reference coupled-cluster algorithm and the calibration of these methods for bond breaking.

Clearly, higher-order excitations (with respect to the Hartree-Fock reference) must be included in an approximate wave function when (quasi)-degeneracies are present. A carefully defined active space can help decrease the number of configurations included in the wave function, but as the system size increases the orbital active spaces become more difficult to define. A general selection scheme, independent of "chemical intuition", could minimize the number of configurations, particularly the more expensive higher-order excitations, included in the coupled-cluster wave function.

Chapter 8 describes a study of the important configurations in configuration interaction and coupled-cluster wave functions, with the goal of constructing compact truncated wave functions to attain chemical accuracy across a potential energy surface.

CHAPTER 2

A THEORETICAL STUDY OF ETHYLENE CATION

2.1 Abstract

The equilibrium geometry, barrier to planarity, and harmonic vibrational frequencies were determined for the ground state of the ethylene cation using several quantum mechanical methods and basis sets. The minimum energy structure is a non-planar D_2 conformer separated from its symmetry equivalent by a planar transition state. The CCSD(T)/cc-pVTZ level of theory obtained an equilibrium C-C bond length and torsion angle of 1.4004 Å and 21°, respectively, which are 0.005 Å and 4.0° less than the experimentally derived values. The documented reliability of CCSD(T)/cc-pVTZ equilibrium geometries may call into question the experimentally derived geometry. In addition, the barrier to planarity has been determined using a series of basis sets and methods aimed at reaching the complete basis set limit. The final vibrationless barrier was determined to be 116 ± 35 cm^{-1}. Also, to aid in the interpretation of a recent infrared cavity ring-down experiment, the harmonic vibrational frequencies were determined at the CCSD(T)/ TZ2P level of theory. After the harmonic frequencies were scaled by a factor accounting for incompleteness in the basis set and electron correlation treatment, the difference between the theoretically and experimentally deduced $\omega_7(b_1)$ frequency was a mere 1.4%.[1]

2.2 Introduction

The values for the torsion angle and barrier to planarity of the ethylene cation have been the topic of controversy for nearly twenty years. This is surprising considering that this molecule is the simplest possible unsaturated hydrocarbon radical cation. The controversy began in 1978 when Köppel and co-workers [8] computed the vibrational structure of the

[1]M. L. Abrams, E. F. Valeev, C. D. Sherrill, and T. D. Crawford, J. Phys. Chem. A 106, 2671-2675 (2002).

first band ($^2B_{3u}$) in the photoelectron spectrum (PES) of neutral ethylene. The Hartree-Fock and many-body methods utilized in the study took into account the vibronic coupling of the $^2B_{3u}$ ground state and $^2B_{3g}$ first excited state of the ion, both in D_{2h} symmetry. After adjusting the energy separation between the $^2B_{3u}$ and $^2B_{3g}$ states, the frequency of the ω_4 vibrational mode, and the vibronic coupling constant to fit the PES of neutral ethylene, they concluded that the ground state was non-planar with a torsion angle and barrier to planarity of approximately 25 and 234 cm^{-1}, respectively.

Prior to the study by Köppel and co-workers, there was no definitive prediction of the torsion angle of the cation. Semiempirical studies [9, 10] predicted a twisted D_2 structure, but disagreed with Hartree-Fock studies [11–13] that predicted a planar D_{2h} structure. However, in 1984 Handy and co-workers [14] found a torsion angle of 20.1° using second order Møller-Plesset perturbation theory (MP2) and a small 3-21G basis. These co-workers also estimated the barrier to planarity to be approximately 100 cm^{-1} using CEPA-1 (coupled electron pair approximation) and a series of Pople basis sets from 3-21G to 6-311G(df,p). More recently, Salhi-Benachenhou co-workers [15] using the quadratic configuration interaction method with single, double, and perturbative triple excitations [QCISD(T)] and a 6-311G(d,p) basis set found a torsion angle of 20.2°. At that geometry, coupled cluster with single, double, and perturbative triple excitations [CCSD(T)] and QCISD(T) with a 6-311++G(2df,p) basis set found the barrier to planarity to be 87 and 104 cm^{-1}, respectively.

Density functional methods have also been used to address the ethylene cation structural dilemma. Eriksson and co-workers [16], using the gradient-corrected Becke exchange and Perdew correlation functional (BP) and a DZP basis, found a torsion angle of 33°, while Liu and co-workers [17], using B3LYP [the Becke three-parameter hybrid exchange functional (B3) with the correlation functional of Lee-Yang-Parr (LYP)] and a 6-311G(d,p) basis, found a torsion angle of 28.4°.

On the experimental front, more recent studies by Toriyama and Okazaki [18, 19] utilized ESR techniques to analyze the hyperfine couplings of ^1H and ^13C. They estimated that the torsion angle of the cation is between 8° and 3°. Therefore, the most recent theoretical

and experimental data suggest a twisted D_2 structure for the ethylene cation with a torsion angle and barrier to planarity as small as 8 and 80 cm^{-1} or as large as 35 and 250 cm^{-1}, respectively.

Several research groups have also analyzed the vibrational spectrum of the cation. From the photoelectron spectrum of neutral ethylene, Pollard and co-workers [20] assigned the $\omega_2(a_g)$ C-C stretching mode and the $\omega_3(a_g)$ H-C-H bending mode of the cation. Somasundram and Handy [21] used scaled Hartree-Fock calculations of neutral ethylene to verify the results of Pollard and co-workers but only for two of the possible twelve modes. Recently, Draves and Taylor [22] utilizing infrared cavity ring-down spectroscopy have observed and assigned the $\omega_7(b_1)$ CH_2 wagging mode.

The goals of the present study are threefold: (1) to determine the equilibrium geometry of the cation using state of the art ab initio techniques; (2) to obtain a definitive estimate for the barrier to planarity using the basis set limit extrapolation techniques suggested by Feller [23, 24] and Helgaker and co-workers [25]; (3) to obtain harmonic vibrational frequencies for comparison to experimental assignments.

2.3 Methods Section

We determined the equilibrium geometry, barrier to planarity, and harmonic vibrational frequencies of the $^2B_3$ ground state ($1a^2\ 1b_1^2\ 2a^2\ 2b_1^2\ 1b_2^2\ 3a^2\ 1b_3^2\ 2b_3^1$) of the ethylene cation using several theoretical methods and basis sets. Geometry optimizations were performed using all-electron B3LYP [26, 27], MP2 [28], CCSD [29], and CCSD(T) [30] methods with spin-unresticted orbitals. Energy gradients were evaluated analytically in all cases [31–33]. The four methods employed three different Gaussian basis sets: DZP, TZ2P, and cc-pVTZ [34]. The DZP basis is the standard double-zeta set of Huzinaga and Dunning augmented with a set of d-functions on carbon and a set of p-functions on hydrogen, denoted C(9s5p1d/4s2p1d) and H(4s1p/2s1p). The TZ2P basis is the standard triple-zeta set of Dunning augmented with two sets of d-functions on carbon and two sets of p-functions on hydrogen, denoted C(10s6p2d/5s3p2d) and H(5s2p/3s2p). The largest basis used in optimizing the geometry is the correlation-consistent polarized valence triple-zeta set of

Dunning and co-workers with the C(10s5p2d1f/4s3p2d1f) and H(5s2p1d/3s2p1d) contraction scheme. Spherical harmonics were employed throughout except for the DZP and TZ2P basis sets.

The barrier to planarity was determined by subtracting the energy of the fully optimized structure of D_2 symmetry from the energy of the planar transition state. The Hartree-Fock contribution to the barrier was determined using Dunning's cc-pVXZ (X = 3, 4, 5) and aug-cc-pVXZ (X = 3, 4, 5) basis sets, where aug- specifies the cc-pVXZ basis augmented by a single diffuse function per angular momentum (1s1p1d1f for X = 3, 1s1p1d1f1g for X = 4, and 1s1p1d1f1g1h for X = 5) [34]. The valence MP2 and valence CCSD(T) contributions to the barrier were determined using Dunning's cc-pVXZ (X = 3, 4) and aug-cc-pVXZ (X = 3, 4) basis sets. The valence CCSD with complete treatment of triple excitations (CCSDT) [35] contribution to the barrier was determined using Dunning's cc-pVXZ (X =2, 3) basis sets. The effect of quadruple excitations was estimated by the Brueckner-reference coupled-cluster method with double excitations (B-CCD) [36]. To isolate the contribution of the quadruple excitations to the barrier, labeled [Q], the total energy of B-CCD(T) [36] was subtracted from the total energy of B-CCD(TQ) [37] at both the twisted minimum and the planar transition state. Due to the high expense of the B-CCD(TQ) calculations, we were restricted to the small cc-pVDZ basis set.

For the valence contribution to the barrier, the 1s core orbitals of the carbon atoms were constrained to be doubly occupied. Core correlation effects were then determined at the MP2 level of theory from all electron treatments with Dunning's cc-pCVXZ (X = 3, 4) basis sets [34]. Harmonic vibrational frequencies were determined using analytic second derivatives [38, 39] for B3LYP/cc-pVTZ and CCSD(T)/(DZP, TZ2P, cc-pVTZ) levels of theory. The Hartree-Fock calculations were performed with PSI 3.0 [40], MP2 and B3LYP calculations were performed with Q-Chem 2.0 [41], CCSD, CCSD(T),and CCSDT calculations were performed with ACESII [42], and B-CCD(T) and B-CCD(TQ) calculations were performed with Gaussian 98 [43].

2.4 Results and Discussion

2.4.1 Geometry analysis

The geometrical parameters of the equilibrium structure of ethylene cation and the barrier to planarity with respect to the constrained planar transition state are presented in Table 1. For all theoretical methods employed in this study, there is a substantial change in the optimized geometry when going from a double-zeta quality basis set with a single set of polarization functions to a triple-zeta quality basis set with two sets of polarization functions: r_{CC} decreases by ~ 0.02 , r_{CH} decreases by ~ 0.01 , θ_{HCC} increases by $\sim 0.1°$, and τ changes as little as $1.6°$ for B3LYP and as much as $10.3°$ for MP2. The only major change in geometry that occurs by adding f-functions to the carbon atoms and d-functions to the hydrogen atoms (i.e. going from the TZ2P basis to the cc-pVTZ basis) is the change in τ, which increases by $2.2°$ at the MP2 level. In general, our results show that r_{CC} and r_{CH} contract, θ_{HCC} is nearly static, and both the torsion angle and the barrier to planarity increase as the basis set is improved and/or as the amount of dynamical correlation energy recovered is increased from MP2 to CCSD to CCSD(T).

Presumably, cc-pVTZ should be the best basis set for both B3LYP and CCSD(T). Though the CCSD(T) geometry approaches that of Köppel and co-workers [8] as size of the basis set is increased, the B3LYP results move further away. The B3LYP/cc-pVTZ geometry contracts r_{CC} by 0.012 Å and overestimates the torsion angle by $7.7°$ with respect to the CCSD(T)/cc-pVTZ geometry. The experimentally derived geometry of Köppel and co-workers [8] is between the B3LYP/cc-pVTZ and CCSD(T)/cc-pVTZ results. The QCISD(T)/6-311G(d,p) results of Salhi-Benachenhou and co-workers [15] compare well with our CCSD(T)/cc-pVTZ results, though r_{CC} differs by 0.01 Å. Overall, the B3LYP/cc-pVTZ and CCSD(T)/cc-pVTZ geometries are similar and exemplify the expected change of removing a π-bonding electron from neutral ethylene (r_{CC} 1.339 Å, r_{CH} 1.086 Å, θ_{HCC} $124.8°$) [44] . The more pressing issue is the magnitude of the barrier to planarity considering the dramatic differences between the experimentally derived barrier, the Salhi-Benachenhou QCISD(T)/6-311++G(2df,p) barrier, and our B3LYP/cc-pVTZ and CCSD(T)/cc-pVTZ values for the barrier.

Table 1: Geometry and barrier to planarity of the ethylene cation.[a,b]

Method/Basis	r_{CC}	r_{CH}	θ_{HCC}	τ	ΔE_e
MP2/DZP	1.4271	1.0906	120.4	0.0	0.0
	1.4271	1.0906	120.4	0.0	
MP2/TZ2P	1.4099	1.0804	120.5	10.3	3.7
	1.4127	1.0800	120.5	0.0	
MP2/cc-pVTZ	1.4049	1.0785	120.5	12.5	7.7
	1.4090	1.0780	120.4	0.0	
CCSD/DZP	1.4256	1.0933	120.5	11.0	5.9
	1.4287	1.0929	120.4	0.0	
CCSD/TZ2P	1.4073	1.0827	120.6	15.2	21.7
	1.4139	1.0819	120.5	0.0	
CCSD/cc-pVTZ	1.4022	1.0804	120.6	16.0	27.0
	1.3940	1.0794	120.5	0.0	
CCSD(T)/DZP	1.4253	1.0950	120.5	16.7	31.6
	1.4316	1.0940	120.4	0.0	
CCSD(T)/TZ2P	1.4058	1.0849	120.6	20.2	69.8
	1.4178	1.0833	120.5	0.0	
CCSD(T)/cc-pVTZ	1.4004	1.0827	120.6	21.0	82.5
	1.4132	1.0809	120.4	0.0	
B3LYP/DZP	1.4079	1.0949	120.9	26.8	271.6
	1.4295	1.0920	120.6	0.0	
B3LYP/TZ2P	1.3891	1.0884	121.1	28.4	354.6
	1.4153	1.0848	120.7	0.0	
B3LYP/cc-pVTZ	1.3880	1.0892	121.2	28.7	378.4
	1.4146	1.0854	120.7	0.0	
QCISD(T)/6-311G(d,p)	1.414	1.092	120.7	20.2	103.7[c]
	1.425	1.091	121.9	0.0	
experimentally derived[d]	1.405	1.091	121.8	25	233.9

[a] Bond lengths are in Å, angles are in degrees, and the barrier is in cm^{-1}. [b] For each level of theory, the first line contains the data for the equilibrium geometry and the second line contains the data for the planar transition state. [c] Salhi-Benachenhou and co-workers [15] obtained the barrier to planarity using QCISD(T)/6-311++G(2df,p) at the the QCISD(T)/6-311G(d,p) geometry. [d] Köppel and co-workers [8]

2.4.2 Barrier to planarity

Although the barrier to planarity is rather small, probably less than 400 cm^{-1}, we have attempted to make as accurate a determination as is possible at this time. Previous theoretical determinations of the barrier utilized standard Pople basis sets, and even though high level theoretical methods [i.e. QCISD(T) and CCSD(T)] were also utilized, no attempt was made to estimate the ab initio limit of the barrier, which can be achieved through schemes

similar to Allen's focal-point analysis [45]. Feller [23, 24] has shown that the lowering of the Hartree-Fock energy exhibits an exponential convergence with respect to the cardinal number X of the Dunning correlation consistent basis sets:

$$E_X = E_{CBS} + ae^{(-bX)} \tag{1}$$

The extrapolated Hartree-Fock barrier obtained by fitting equation 1 to the cc-pV(3,4,5)Z HF energies is -291.4 cm^{-1} (see Table 2). The difference between the cc-pV(3,4,5)Z and aug-cc-pV(3,4,5)Z extrapolated Hartree-Fock barriers is only 0.3 cm^{-1}, thus showing that the Hartree-Fock basis set limit has confidently been converged to within 1 cm^{-1}. The extrapolation of the correlation energy uses a two-point inverse power scheme proposed by Helgaker and co-worker [25]:

$$E_{CBS}(X,Y) = \frac{E^X X^3 - E^Y Y^3}{X^3 - Y^3} \tag{2}$$

where E^X and E^Y denote the correlation energies obtained from correlation consistent basis sets with cardinal numbers X and Y. The cc-pV(3,4)Z MP2 contribution is estimated at +279 cm^{-1}, still not enough to stabilize the twisted structure. The addition of diffuse functions stabilizes the planar transition state by 11 cm^{-1}. The bulk of the total electronic barrier is due to higher-order electron correlation effects. The CCSD and CCSD(T) results do show a stabilizing effect for the twisted conformation, contributing an additional +29 and +71 cm^{-1}, respectively. The CCSDT results show a +19 cm^{-1} contribution, still a considerable portion of the total. The correlation energy from quadruple excitations [Q] contributes only -2 cm^{-1}, thus suggesting that quadruple and higher excitations contribute very little to the barrier.

Table 2: Valence focal-point analysis of the barrier to planarity (in cm^{-1}) of ethylene cation. [a]

	ΔE_e(UHF)	δMP2	δCCSD	δCCSD(T)	δCCSDT	δ[Q][b]	ΔE_e(Total)
cc-pVDZ (76)[c]	-281	+219	+32	+58	+18	-2	+44
cc-pVTZ (116)	-290	+256	+29	+67	+19	[-2]	[+61]
cc-pCVTZ (142)	-290	+273					
aug-cc-pVTZ (182)	-283	+277	+29	+69	[+19]	[-2]	[+92]
cc-pVQZ (230)	-291	+270	+29	+71	[+19]	[-2]	[+78]
cc-pVCQZ (288)	-290	+286					
aug-cc-pVQZ (344)	-289	+272	[+29]	[+71]	[+19]	[-2]	[+100]
cc-pV5Z (402)	-291		[+29]	[+71]	[+19]	[-2]	[+100]
aug-cc-pV5Z (574)	-291		[+29]	[+71]	[+19]	[-2]	[+100]
extrapolate limit (∞)	[-291]	[+274]	[+29]	[+71]	[+19]	[-2]	[+100]

[a] The values in brackets under ΔE_e(UHF) is the average of the barrier obtained by fitting eq 1 to the cc-pV(3,4,5)Z and aug-cc-pV(3,4,5)Z Hartree-Fock energies. The value in brackets under δMP2 is the average of the barrier obtained by fitting eq 2 to the cc-pV(3,4)Z and aug-cc-pV(3,4) MP2 energies. All the values in brackets under the coupled-cluster columns were obtained by assuming that the basis set is saturated for the given correction. [b] The contribution of the quadruple excitations to the barrier is determined by δBCCD(TQ) - δBCCD(T). [c] Numbers in parentheses are the numbers of contracted Gaussian functions in the given basis sets.

The core correlation contribution was determined by subtracting the the frozen-core MP2/cc-pCVQZ contribution from the all-electron MP2/cc-pCVQZ contribution, which yielded a core correlation contribution of +16 cm^{-1}. To reach a final estimate of the barrier to planarity, the valence contribution is added to the core correlation contribution. We also include the possible error in the magnitude of the barrier to planarity due to basis set truncation and neglect of higher-order correlation. The basis set truncation error was estimated by the difference between contributions to the barrier from the cc-pVXZ and aug-cc-pVXZ series of basis sets, while the neglect of higher-order correlation was estimated by the reduction in the contribution to the barrier from the UHF→MP2→CCSD→CCSD(T) →CCSDT→[Q] series. Given the small contribution from [Q] we estimate that the effect of higher-order correlation contributions is probably less than 10 cm^{-1}. The final vibrationless barrier to planarity, without relativistic or non-Born-Oppenheimer corrections, is (100 + 16) ± (25 + 10) = 116 ± 35 cm^{-1}.

2.4.3 Vibrational analysis

The calculated harmonic vibrational frequencies of the cation are reported in Table 3. Our primary emphasis is on the second, third, fourth, and seventh modes since we have experimental and theoretical data for comparison. The results of previous theoretical determinations of the harmonic vibrational frequencies started from a D_{2h} reference geometry, though the cation is now established to have D_2 symmetry at equilibrium. Another problem encountered in previous theoretical studies was that the highest level of theory used was the Hartree-Fock method. Hartree-Fock theory is known to overestimate the vibrational frequencies, though Somasundram and Handy [21] scaled their results with $(\omega_{expt}/\omega_{calc})$ for the neutral to compensate for basis set incompleteness and neglect of higher-order correlation effects.

Table 3: Harmonic vibrational frequencies (in cm^{-1}) of the ethylene cation

Mode	B3LYP/cc-pVTZ	CCSD(T)/DZP	CCSD(T)/TZ2P	CCSD(T)/cc-pVTZ	Other
ω_1(a)	3082 (0)	3168 (0)	3133 (0)	3149 (0)	
ω_2(a)	1502 (0)	1563 (0)	1550 (0)	1557 (0)	1510[b]
ω_3(a)	1273 (0)	1275 (0)	1269 (0)	1290 (0)	1264[b]
ω_4(a)[c]	616 (0)	328 (0)	402 (0)	420 (0)	887[d]
ω_5(b_1)	3176 (3)	3284 (1)	3243 (2)	3229 (2)	
ω_6(b_1)	1206 (11)	1229 (7)	1240 (11)	1237 (13)	
ω_7(b_1)	887 (37)	944 (74)	943 (69)	955 (70)	908[e]
ω_8(b_2)	3196 (84)	3300 (80)	3261 (108)	3249 (110)	
ω_9(b_2)	1078 (1)	1069 (0)	1110 (1)	1122 (1)	
ω_10(b_2)	785 (5)	819 (7)	822 (8)	815 (9)	
ω_11(b_3)	3086 (87)	3160 (69)	3129 (90)	3143 (91)	
ω_12(b_3)	1432 (55)	1467 (62)	1470 (71)	1472 (76)	

[a] Values in parentheses are the infrared intensities in km mol^{-1}. [b] Experimentally derived harmonic vibrational frequencies from Pollard and co-workers [20] [c] The ω_4 mode is highly anharmonic and should not be estimated by a harmonic potential. [d] Köppel and co-workers [8] [e] Draves and Taylor [22]

The four levels of theory used for determining the harmonic vibrational frequencies are, B3LYP/cc-pVTZ, CCSD(T)/DZP, CCSD(T)/TZ2P, CCSD(T)/cc-pVTZ, all at their respectively optimized geometries. The CCSD(T)/(DZP, TZ2P, cc-pVTZ) level of theory was chosen for two reasons. First, the computational cost of determining the harmonic vibrational frequencies with the CCSD(T) method and a larger basis set is beyond our reach. Second, Thomas and co-workers [46, 47] performed a systematic study on the accuracy of CCSD(T) with these three basis sets. These studies provided an approximate scaling factor of (2.4, 2.3, 2.2)% for the CCSD(T)/(DZP, TZ2P, cc-pVTZ) harmonic vibrational frequencies, respectively. The density functional results are included for comparison to the CCSD(T) results. The harmonic frequency of the C-C stretching mode, ω_2(a), was experimentally derived by Pollard and co-workers [20]. All four levels of theory compare very well to the experimentally derived harmonic frequency with unscaled values varying by (-8, +53, +40, +47) cm^{-1} and the scaled CCSD(T) values varying by (+15, +4, +13) cm^{-1} from experiment. The harmonic frequency of the H-C-H bending mode, ω_3(a), also agrees very well with the experimentally derived harmonic frequency with unscaled values varying by (+9, +11, +5, +26) cm^{-1} and the scaled CCSD(T) values varying by (-20, -24, -2) cm^{-1} from experiment.

The fourth vibrational mode, ω_4(a), is a torsion mode and is highly sensitive to the optimized torsion angle as well as the level of theory. Pollard and co-workers [20] did measure the $2\nu_4 = 441\pm 4$ cm^{-1} transition and noted that the (0001) transition is too weak to be observed in the 21.2 eV PES, even with observed resolution of 12-13 meV FWHM. The harmonic frequency of the ω_4(a) mode is more than five times the height of the barrier to planarity at the CCSD(T)/cc-pVTZ level of theory. We followed the procedure of Valeev and co-workers [48] for determining the bound eigenstates of a one-dimensional potential: the Mathematica package [49] was used to evaluate the G(ν,ν) matrix element along the torsional path, to construct an interpolating function for the CCSD(T)/cc-pVTZ potential, and to apply the Cooley-Numerov procedure for determining the exact torsional eigenstates, though the splitting of energy levels due to the $\sim$ 10,500 cm^{-1} barrier at τ = 90° was neglected. The first five energy levels were determined to be: ϵ_τ = (122, 324,

688, 1093,1545) cm^{-1}. Assuming the torsion mode does not couple to any other vibrational modes, the fundamental transition and the first three overtones are $G(\nu_4)$ = (202, 566, 971, 1423) cm^{-1} compared to the experimentally derived values of $G(\nu_4)$ = ([not observed], 438, 766, 1158) cm^{-1}. At the CCSD(T)/cc-pVTZ level of theory, the ground vibrational state is above the barrier to planarity (83 cm^{-1}), thus one has to consider the possibility that the ethylene cation is a quasiplanar molecule with a large amplitude torsional motion. However, as the previous section of the paper has shown, the estimated ab initio limit of the barrier to planarity is 116 $\pm$ 35 cm^{-1}. Therefore, to reach a definitive conclusion as to the nature of the lowest vibrational energy level due to the low energy barrier, the torsion potential must be determined using methods approaching the ab initio limit, which is currently beyond our capability. Köppel and co-workers used 887 cm^{-1} for ω_4 (see Table 3) in their empirical vibronic coupling model; however, our results show that the fundamental transition of the fourth vibrational mode is approximately 200 cm^{-1}.

Our motivation for determining the harmonic vibrational frequencies of the cation was to aid in the assignment of the infrared cavity ring-down spectrum of Draves and Taylor [22]. Their results showed a two-peak signature that was duly assigned to the $\omega_7(b_1)$ CH_2 wag. The second peak was assigned as the first overtone with an anharmonicity ($\omega_e x_e$) of $\sim$ 9 cm^{-1}. From their analysis, they determined that the harmonic $\omega_7(b_1)$ frequency is $\sim$ 908 cm^{-1}. Using the unscaled and scaled CCSD(T)/(DZP, TZ2P, cc-pVTZ) result, theory and experiment are in agreement within (3.8, 3.7, 4.9)% and (1.4, 1.4, 2.8)% and B3LYP/cc-pVTZ is underestimated by 2.4%.

2.5 Conclusions

The following conclusions can be drawn from the results presented in this study: (1) The CCSD(T)/cc-pVTZ equilibrium geometry compares well with the results of Salhi-Benachenhou and co-workers [15], but contracts r_{CC} by 0.005 Å and τ decreased by 4.0° compared to Köppel and co-workers's experimentally derived values [8] , (2) The barrier to planarity has been estimated using a series of methods and basis sets designed to extrapolate to the ab initio limit. The final vibrationless barrier of 116 $\pm$ 35 cm^{-1} is the result

of the most systematic and comprehensive examination of this barrier to date, the B3LYP harmonic vibrational frequencies are more accurate than the unscaled CCSD(T) harmonic vibrational frequencies, though both the B3LYP and the scaled CCSD(T) harmonic vibrational frequencies compare very well with the experimental results of Pollard and Draves and Taylor, exact torsion vibrational energy levels evaluated at the CCSD(T)/cc-pVTZ level of theory, (3) neglecting mode-coupling, all lie above the barrier to planarity. A definitive conclusion of the torsion vibrational mode will require a torsion potential approaching the ab initio limit, including mode-coupling, which is currently beyond our capability

CHAPTER 3

FULL CONFIGURATION INTERACTION BENCHMARKS: I

3.1 Abstract

We compare several standard polarized double-zeta basis sets for use in full configuration interaction benchmark computations. The 6-31G**, DZP, cc-pVDZ, and Widmark-Malmqvist-Roos atomic natural orbital (ANO) basis sets are assessed on the basis of their ability to provide accurate full configuration interaction spectroscopic constants for several small molecules. Even though highly correlated methods work best with larger basis sets, predicted spectroscopic constants are in good agreement with experiment; bond lengths and harmonic vibrational frequencies have average absolute errors no larger than 0.015 Å and 1.6%, respectively, for all but the ANO basis. For the molecules considered, 6-31G** gives the smallest average errors, while the ANO basis set gives the largest. The use of variationally optimized basis sets and natural orbitals are also explored for improved benchmarking. Although optimized basis sets do not always improve predictions of molecular properties, taking a DZP-sized subset of the natural orbitals from a singles and doubles configuration interaction computation in a larger basis significantly improves results.[1]

3.2 Introduction

Wave function based quantum chemical methods approach the exact solution to the electronic Schrödinger equation as the basis set and the treatment of electron correlation are simultaneously improved. In recent years, considerable progress has been made in understanding convergence toward the complete basis set limit [50–56]. It has been more

[1]M. L. Abrams and C. D. Sherrill, J. Chem. Phys. 118, 1604-1609 (2003).

challenging to investigate convergence of electron correlation because full configuration interaction (FCI), which provides the exact solution of the electronic Schrödinger equation within a given basis set, has a computational cost that increases factorially with the number of electrons or orbitals.

Fortunately, both algorithmic advances [57–64] and improvements in computer hardware have made FCI benchmarks less computationally expensive. Whereas FCI benchmarking in the 1980s and early 1990s focused almost exclusively on single-point energies, it is now possible to perform geometry optimizations and even frequency analysis for very small molecules using FCI to examine the effect of higher-order correlation on molecular properties [65–73]. It is also now possible to afford enough FCI computations to generate potential energy curves, [74–77] which are beneficial for assessing the reliability of standard quantum chemistry methods for bond-breaking reactions.

In light of these expanded possibilities, it is important to ask which basis sets are best for FCI benchmarking. Since FCI computations with basis sets larger than polarized double-zeta are rarely possible, we compare several standard basis sets of this size to determine their suitability for the computation of FCI molecular properties. Specifically, Dunning's DZP [78] and cc-pVDZ [50] basis sets, Pople's 6-31G** basis,[79] and the Widmark-Malmqvist-Roos atomic natural orbital (WMR ANO) [80] polarized double zeta basis sets are used to compute spectroscopic constants for the ground states of the BH, CH^+, NH, OH^+, HF, and C_2 molecules.

The use of variationally optimized scale factors and natural orbitals are examined for their ability to provide improved one-particle spaces for the FCI. Natural orbitals (NOs) are those orbitals which diagonalize the one-particle density matrix; they provide the most rapidly convergent CI expansion in the sense that to achieve a given accuracy requires fewer configurations in a natural orbital basis than in any other orthonormal basis for a given underlying one-particle space [81]. The NOs are ordered according to their occupation numbers (one-particle density matrix eigenvalues), and larger occupation numbers reflect larger contributions to the CI wave function from which the orbitals are derived. Hence, one might expect that the most important m orbitals from a natural orbital computation

should provide a compact orbital subspace for a subsequent FCI computation. Here we investigate the efficiency of performing CI singles and doubles (CISD) computations in the large cc-pVQZ basis set and taking the m most important NOs, where m is equal to the number of orbitals present in a polarized double-zeta basis. This procedure gives far better results than any of the standard basis sets considered.

3.3 Theoretical Approach

Several spectroscopic constants and dissociation energies of the ground states of BH, CH^+, NH, OH^+, HF, and C_2 were determined using the FCI method as implemented in the DETCI program [82] in PSI 3.0 [40].

The customary rotational and vibrational energy level expressions for a diatomic molecule are

$$F_v(J) = B_v J(J+1) - \overline{D}_v[J(J+1)]^2 \cdots \tag{3}$$

$$G(v) = \omega_e \left(v + \frac{1}{2}\right) - \omega_e x_e \left(v + \frac{1}{2}\right)^2 \cdots \tag{4}$$

with B_v and $\overline{D}_v$ defined as

$$B_v = B_e - \alpha_e \left(v + \frac{1}{2}\right) \cdots \tag{5}$$

$$\overline{D}_v = \overline{D}_e + \beta_e \left(v + \frac{1}{2}\right) + \cdots \tag{6}$$

We have determined the harmonic vibrational frequency, ω_e, anharmonicity constant, $\omega_e x_e$, rotational constant, B_e, vibration-rotation coupling constant, α_e, and the centrifugal distortion constant, $\overline{D}_e$, using second through fourth-order force constants, $f_{rr} - f_{rrrr}$. The force constants were computed using tightly converged energies ($\sim 10^{-10}$ hartrees) from 5 geometries equally distributed around the equilibrium bond distance. The dissociation energies were computed using the supermolecule approach for all cases.[2]

Computations were performed using five standard basis sets. The 6-31G** basis set is contracted as (10s4p1d/3s2p1d) and (4s1p/2s1p) for first-row elements and hydrogen,

[2]CISD is not size consistent, so the quality of CISD natural orbitals will degrade at extended geometries. Computing dissociation energies by a fragment approach makes predictions even better on average; the difference ranges from 0.04 eV for BH to 0.32 eV for HF. The performance of CISD natural orbitals beyond equilibrium remains to be fully explored.

respectively [79]. The Huzinaga-Dunning DZP basis set [78, 83] is contracted as (9s 5p 1d / 4s 2p 1d) and (4s 1p / 2s 1p), and Dunning's correlation consistent cc-pVDZ basis set [50] is contracted as (9s4p1d/3s2p1d) and (4s1p/2s1p). The WMR-ANO basis set is contracted as (14s9p4d/3s2p1d) and (8s4p/2s1p) [80]. The five basis sets use a different number of d functions: 6-31G**, DZP, and WMR-ANO use six cartesian d functions, while it is customary to use five pure angular momentum d functions for cc-pVDZ. The effect of varying the number of d polarization functions was examined by using cc-pVDZ with 6 cartesian d functions and 6-31G** with 5 pure angular momentum d functions, designated cc-pVDZ(6d) and 6-31G**(5d), respectively.

To investigate the possibility of improving results without increasing the size of the basis, a modified 6-31G** basis for BH was constructed by optimizing a scale factor for each atom, i.e. the primitive gaussian exponents are multiplied by the square of the scale factor (excluding the core function). This basis set is denoted 6-31G**(opt). In an alternative strategy, separate CISD/cc-pVQZ natural orbital computations were performed at each geometry and the NOs with the largest occupation numbers were retained to form a set of molecular orbitals, denoted DZP-NO, having the same number of orbitals as cc-pVDZ (the remaining weakly occupied NOs were discarded). Core orbitals were constrained to be doubly-occupied in all computations, and for the DZP basis the corresponding high-lying virtual orbitals were deleted.

3.4 Results and Discussion

Total electronic energies and spectroscopic constants are presented in Tables 4 (BH), 5 (CH^+), 6 (NH), 7 (OH^+), 8 (HF), and 9 (C_2). Experimental results are taken from Huber and Herzberg [84] except for CH^+, where results are from Carrington and Ramsey [85] (the Huber and Herzberg value for ω_e of CH^+, cited in many theoretical works, is more than 100 cm^{-1} too low). Experimental dissociation energies D_0 have been converted to D_e using the experimental ω_e and $\omega_e x_e$ for a more direct comparison to the theoretically computed results. Table 10 presents average absolute errors for the set of molecules considered.

3.4.1 Comparison of standard basis sets

In general, improved descriptions of electron correlation result in larger predicted bond lengths, while larger basis sets result in shorter predicted bond lengths. Systematic studies have shown that highly correlated methods such as CCSD(T) can require large basis sets (often triple or quadruple zeta) to provide very accurate predictions of bond lengths[86, 87], although CCSD(T) harmonic vibrational frequencies are predicted surprisingly well by the modest DZP or cc-pVDZ basis sets [46]. One might expect that FCI, representing the complete treatment of electron correlation for a given basis set, might require even larger basis sets. However, the present FCI spectroscopic constants are generally in good agreement with experiment. Bond lengths are systematically overestimated, with the most severe overestimates occurring for the WMR ANO basis (errors of about 0.03 Å for BH, CH^+, and C_2). The average absolute errors (cf. Table 10) are around 0.01 - 0.02 Å for the standard basis sets except for 6-31G**, which gives lower errors of 0.006 Å. This compares to an average overestimation of 0.02 Å for single bonds in several small molecules at the CCSD(T)/cc-pVDZ level of theory reported by Martin [86]; hence, the basis set requirements of FCI do not seem significantly different from those of CCSD(T). The rotational constant B_e, which depends on the equilibrium bond length, is likewise predicted most accurately for 6-31G** (1.2% error) and least accurately for WMR-ANO (4.3% error).

Table 4: Spectroscopic constants of $X\ ^1\Sigma^+$ BH.[a]

Method	Energy	r_e	ω_e	$\omega_e x_e$	B_e	$\overline{D}_e$	α_e	D_e
FCI/DZP	-25.208881	1.2491	2339	47.6	11.700	0.12e-02	0.399	3.48
FCI/6-31G**(5d)	-25.206493	1.2346	2392	52.2	11.977	0.12e-02	0.429	3.51
FCI/6-31G**	-25.207157	1.2344	2388	52.0	11.980	0.12e-02	0.431	3.51
FCI/cc-pVDZ	-25.215324	1.2559	2340	48.8	11.574	0.11e-02	0.396	3.44
FCI/cc-pVDZ(6d)	-25.216182	1.2553	2343	49.0	11.584	0.11e-02	0.398	3.44
FCI/WMR-ANO	-25.214743	1.2675	2309	49.5	11.364	0.11e-02	0.388	3.47
FCI/6-31G**(opt)	-25.211093	1.2446	2376	52.5	11.786	0.12e-02	0.415	3.49
FCI/DZP-NO	-25.229338	1.2366	2354	50.9	11.940	0.12e-02	0.457	3.57
FCI/DZP-NO(5Z)	-25.229704	1.2362	2350	51.8	11.948	0.12e-02	0.446	3.57
Experiment		1.2324	2367	49.4	12.021	0.12e-02	0.412	3.57

[a] Energy is in a.u., r_e in Å, D_e in eV, and all other quantities in cm^{-1}.

Table 5: Spectroscopic constants of $X\ ^1\Sigma^+$ CH^+.[a,b]

Method	Energy	r_e	ω_e	$\omega_e x_e$	B_e	$\overline{D}_e$	α_e	D_e
FCI/DZP	-38.004943	1.1365	2923	66.2	14.038	1.29e-03	0.503	4.02
FCI/6-31G**(5d)	-37.998840	1.1246	2939	66.9	14.334	1.36e-03	0.532	3.97
FCI/6-31G**	-38.000644	1.1241	2939	67.1	14.350	1.37e-03	0.537	3.98
FCI/cc-pVDZ	-38.002366	1.1460	2892	64.6	13.807	1.26e-03	0.492	3.97
FCI/cc-pVDZ(6d)	-38.003779	1.1449	2898	65.9	13.843	1.26e-03	0.497	3.94
FCI/WMR-ANO	-38.002902	1.1623	2805	58.8	13.421	1.23e-03	0.457	3.94
FCI/DZP-NO	-38.023851	1.1319	2840	59.3	14.153	1.41e-03	0.495	4.15
Experiment		1.1309	2858	59.3	14.176	1.37e-03	0.493	4.27

[a] Energies in a.u., r_e in Å, D_e in eV, and all other quantities in cm^{-1}. [b] Experimental data from Carrington and Ramsey [85].

Table 6: Spectroscopic constants of $X\ ^3\Sigma^-$ NH.[a]

Method	Energy	r_e	ω_e	$\omega_e x_e$	B_e	$\overline{D}_e$	α_e	D_e
FCI/DZP	-25.208881	1.2491	2339	47.6	11.700	0.12e-02	0.399	3.48
FCI/6-31G**(5d)	-25.206493	1.2346	2392	52.2	11.977	0.12e-02	0.429	3.51
FCI/6-31G**	-25.207157	1.2344	2388	52.0	11.980	0.12e-02	0.431	3.51
FCI/cc-pVDZ	-25.215324	1.2559	2340	48.8	11.574	0.11e-02	0.396	3.44
FCI/cc-pVDZ(6d)	-25.216182	1.2553	2343	49.0	11.584	0.11e-02	0.398	3.44
FCI/WMR-ANO	-25.214743	1.2675	2309	49.5	11.364	0.11e-02	0.388	3.47
FCI/6-31G**(opt)	-25.211093	1.2446	2376	52.5	11.786	0.12e-02	0.415	3.49
FCI/DZP-NO	-25.229338	1.2366	2354	50.9	11.940	0.12e-02	0.457	3.57
FCI/DZP-NO(5Z)	-25.229704	1.2362	2350	51.8	11.948	0.12e-02	0.446	3.57
Experiment		1.2324	2367	49.4	12.021	0.12e-02	0.412	3.57

[a] Energies in a.u., r_e in Å, D_e in eV, and all other quantities in cm^{-1}.

Although most quantum chemical methods generally overestimate harmonic vibrational frequencies, highly correlated methods such as CCSD(T) occasionally underestimate them [86]. Here, too, we find that several FCI harmonic vibrational frequencies underestimate experiment. The WMR ANO basis, which gives the largest overestimations of bond lengths, usually provides the lowest predicted vibrational frequencies. The average absolute errors for harmonic frequencies are close to 1-2% for any of the standard basis sets considered, which is again similar to the average error in CCSD(T)/cc-pVDZ harmonic frequencies reported by Martin [86]. Vibrational anharmonicities are predicted with roughly similar accuracies for all of the standard basis sets (4-7%), although cc-pVDZ is best on average. Likewise,

Table 7: Spectroscopic constants of $X\ ^3\Sigma^-$ OH^+.[a]

Method	Energy	r_e	ω_e	$\omega_e x_e$	B_e	$\overline{D}_e$	α_e	D_e
FCI/DZP	-75.113655	1.0347	3173	89.8	16.606	0.18e-02	0.763	4.76
FCI/6-31G**(5d)	-75.092302	1.0323	3165	87.8	16.682	0.19e-02	0.757	4.80
FCI/6-31G**	-75.094766	1.0323	3161	86.1	16.684	0.19e-02	0.756	4.80
FCI/cc-pVDZ	-75.110738	1.0383	3123	81.8	16.492	0.18e-02	0.719	4.73
FCI/cc-pVDZ(6d)	-75.112752	1.0377	3124	81.1	16.511	0.18e-02	0.718	4.74
FCI/WMR-ANO	-75.124602	1.0397	3103	89.9	16.448	0.18e-02	0.769	4.66
FCI/DZP-NO	-75.157488	1.0287	3121	83.0	16.804	0.19e-02	0.746	5.10
Experiment		1.0289	3113	78.5	16.794	0.19e-02	0.749	5.29

[a] Energies in a.u., r_e in Å, D_e in eV, and all other quantities in cm^{-1}.

Table 8: Spectroscopic constants of $X\ ^1\Sigma^+$ HF.[a]

Method	Energy	r_e	ω_e	$\omega_e x_e$	B_e	$\overline{D}_e$	α_e	D_e
FCI/DZP	-100.242690	0.9243	4173	94.7	20.618	2.01e-03	0.794	5.74
FCI/6-31G**(5d)	-100.199160	0.9213	4171	97.2	20.750	2.05e-03	0.822	5.55
FCI/6-31G**	-100.201597	0.9214	4172	95.0	20.746	2.05e-03	0.813	5.56
FCI/cc-pVDZ	-100.228652	0.9202	4144	92.8	20.799	2.10e-03	0.806	5.49
FCI/cc-pVDZ(6d)	-100.231198	0.9200	4147	92.7	20.810	2.10e-03	0.804	5.50
FCI/WMR-ANO	-100.277950	0.9286	4068	92.5	20.428	2.06e-03	0.797	5.81
FCI/DZP-NO	-100.308531	0.9189	4126	90.5	20.862	2.13e-03	0.791	5.74
Experiment		0.9168	4138	89.9	20.956	2.15e-03	0.798	6.13

[a] Energies in a.u., r_e in Å, D_e in eV, and all other quantities in cm^{-1}.

predictions of vibration-rotation interaction constants α_e are of comparable quality across the standard basis sets (3-5% errors). These results compare to average absolute errors of 6-9% ($\omega_e x_e$) and 5-8% (α_e) for CCSD(T) with polarized double-zeta basis sets for a few diatomic molecules [88]. Centrifugal distortion constants, depending on B_e and ω_e, are predicted best by 6-31G** and with roughly similar accuracy among the other standard basis sets.

Dissociation energies are more challenging to compute because accurate estimates can require very large basis sets. For example, in N_2, *i*-type polarization functions were found to contribute 0.4 kcal/mol to the dissociation energy [89]. For the standard polarized double-zeta basis sets considered, the errors in D_e are generally within 0.5 eV except for HF and C_2, and the average absolute errors are 8-10%. The 6-31G** and DZP basis sets perform

Table 9: Spectroscopic constants of $X\ ^1\Sigma_g^+$ C_2.

Method	Energy	r_e	ω_e	$\omega_e x_e$	B_e	$\overline{D}_e$	α_e	D_e
FCI/DZP	-75.731641	1.2695	1813	13.3	1.743	6.44e-06	0.017	5.74
FCI/6-31G**(5d)	-75.723459	1.2603	1859	13.1	1.769	6.40e-06	0.017	6.00
FCI/6-31G**	-75.726127	1.2596	1859	13.2	1.771	6.43e-06	0.017	5.99
FCI/cc-pVDZ	-75.729852	1.2727	1813	13.5	1.734	6.35e-06	0.017	5.67
FCI/cc-pVDZ(6d)	-75.732244	1.2717	1814	13.6	1.737	6.37e-06	0.017	5.69
FCI/WMR-ANO	-75.732305	1.2793	1766	15.8	1.716	6.49e-06	0.018	5.44
FCI/DZP-NO	-75.766448	1.2500	1834	12.8	1.797	6.90e-06	0.017	5.99
Experiment		1.2425	1855	13.3	1.820	6.92e-06	0.018	6.33

[a] Energies in a.u., r_e in Å, D_e in eV, and all other quantities in cm^{-1}.

Table 10: Absolute average error of the molecular test set.[a]

Method	r_e	ω_e	$\omega_e x_e$	B_e	$\overline{D}_e$	α_e	D_e
FCI/DZP	0.011	1.6	7.3	2.2	5.1	2.6	7.7
FCI/6-31G**(5d)	0.006	1.2	8.0	1.3	2.1	4.4	7.8
FCI/6-31G**	0.006	1.1	7.3	1.2	2.0	4.6	7.7
FCI/cc-pVDZ	0.015	1.3	3.9	2.9	6.3	2.6	9.6
FCI/cc-pVDZ(6d)	0.014	1.3	4.2	2.8	6.3	2.8	9.6
FCI/WMR-ANO	0.022	2.2	6.9	4.3	6.7	3.1	9.0
FCI/DZP-NO	0.003	0.7	2.7	0.6	0.7	3.4	4.5

[a] Error in bond lengths are in Å and all other quantities are in %.

slightly better on average. The FCI results underestimate D_e because the polarized double-zeta basis sets are not sufficient to accurately describe dynamic electron correlation around equilibrium.

Since the basis set requirements for CI singles and doubles (CISD) are less severe than for FCI, it is possible for CISD to be more accurate than FCI when a polarized double-zeta basis set is used. For the CISD/cc-pVDZ level of theory, the average absolute errors in spectroscopic constants of the first row hydrides [54] are 0.013 Å(r_e), 0.7% (ω_e), 3.0% ($\omega_e x_e$), 2.2% (B_e), 3.7% (α_e), and 11% (D_e). These results are comparable to and in some cases slightly better than our present FCI results due to a cancellation of errors between the CISD approximation and the limited cc-pVDZ basis set. This error cancellation cannot happen for FCI, since the treatment of electron correlation is exact. Fortunately, however, this lack of error cancellation between FCI and polarized double-zeta basis sets does not

lead to large errors in predicted spectroscopic constants. This suggests that basis sets of this size are already large enough for meaningful benchmark studies on the effects of higher levels of electron correlation on molecular properties.

In the present study, 6*d* functions were used for the 6-31G**, DZP, and WMR-ANO basis sets. One might suppose that the additional *d* function could offer a slight advantage to these basis sets. This possibility was explored by performing additional computations with the cc-pVDZ basis using 6*d* functions, denoted cc-pVDZ(6d); spectroscopic constants changed very little compared to 5*d* functions, with bond lengths shortening by about 0.01 Å or less and frequencies changing by just a few cm^{-1}. This suggests that 6*d* functions offer very little advantage over 5*d*, while the computational cost of adding an additional *d* function per heavy atom will increase the cost of the FCI significantly. The differences between 6-31G**(5d) and 6-31G**(6d) are also insignificant, thus supporting the idea of using pure angular momentum *d* functions if limited by computational resources.

Overall, the average absolute errors in Table 10 show that FCI properties computed with the 6-31G** basis are generally better than for the other standard basis sets considered, while the WMR ANO basis gives some of the largest errors compared to experiment. However, except for bond lengths, there is not a large difference among the basis sets considered.

3.4.2 Variationally optimized basis sets

We investigated the possibility of obtaining better results using optimized basis sets. For BH, the 6-31G** basis was modified by optimizing the scale factors for the boron and hydrogen valence functions to give the lowest FCI energy at R(B–H)=1.20 Å. The scale factors thus obtained were 0.93 (B) and 0.92 (H), and the optimized basis sets multiplied the primitive Gaussian exponents by the square of the scale factor except for the primitives comprising the B 1s function. The resulting spectroscopic constants in Table 4, labeled 6-31G**(opt), are improved in some cases, but worse for others compared to the regular 6-31G** basis. The equilibrium bond length prediction is made worse by 0.01 Å.

3.4.3 Natural orbitals

As discussed in the introduction, natural orbitals can provide a very compact orbital subspace for use in CI computations. Here we obtained CISD/cc-pVQZ natural orbitals and deleted the most weakly occupied NOs to achieve the same number of orbitals as in a polarized double-zeta basis set like cc-pVDZ. Full CI computations were then performed in this truncated natural orbital set denoted DZP-NO.

The equilibrium FCI energies are much lower with the DZP-NO orbitals than for any of the standard basis sets, and the spectroscopic constants are much more reliable on average. The average absolute error in r_e drops to a mere 0.003 Å, while for ω_e it is reduced to 0.7%. Similarly, B_e, $\overline{D}_e$, and the dissociation energy are all improved substantially. The quality of the DZP-NO results does not appear to improve significantly by using a larger basis to generate the natural orbitals. For BH, the natural orbitals were also obtained from the cc-pV5Z basis, yielding a set labeled DZP-NO(5Z) in Table 4; the results are very similar to the cc-pVQZ generated DZP-NO values.

3.5 Conclusions

Several standard polarized double-zeta basis sets have been compared for their suitability in full configuration interaction benchmarking by determining their reliability for the spectroscopic constants of several diatomics. The performance of the basis sets is similar, but 6-31G** is better on average. Although FCI might be expected to have very large basis set requirements, the predicted spectroscopic constants are in good agreement with experiment and exhibit errors similar to those of CCSD(T) and not much worse than CISD with similar basis sets. This suggests that the effects on molecular properties of electron correlation beyond CCSD(T) can be reasonably examined in DZP-sized basis sets. The optimization of basis scaling factors did not significantly improve spectroscopic constants for BH. The use of DZP-sized sets of natural orbitals gave results far superior to those of the standard basis sets and may provide considerably more rapid convergence to the complete basis set limit for highly correlated wave functions.

CHAPTER 4

FULL CONFIGURATION INTERACTION BENCHMARKS: II

4.1 Abstract

Complete-active-space self-consistent field (CASSCF), complete-active-space second-order perturbation theory (CASPT2), and two restricted active-space variants of multi-reference configuration interaction (singles, doubles, and limited triples and quadruples, or CISD[TQ], and second-order configuration interaction, or SOCI), have been assessed for bond breaking in BH, HF, and CH_4 by comparison to the full configuration interaction limit. These results allow one, for the first time, to ascertain typical errors for such reactions across the entire potential energy curve. They also provide an assessment of different prescriptions for choosing an active space. A valence active space and a one-to-one active space were considered along with the basis sets cc-pVQZ, 6-31G**, and 6-31G* for BH, HF, and CH_4, respectively. The valence active space performs better than the one-to-one active space for BH, but is inferior for HF. Always choosing the larger of the two active spaces for a given molecule leads to the best results. When using the larger of the two active spaces, the non-parallelity errors for CASPT2, CISD[TQ], and SOCI were less than 3.3, 1.4, and 0.3 kcal/mol, respectively. These results are superior to those of unrestricted coupled-cluster with perturbative triples [UCCSD(T)] for these same systems.[1]

4.2 Introduction

The last decade has seen a major advancement in electron correlation methods based on a multi-configuration wave function, i.e., multi-reference versions of configuration interaction (MRCI), perturbation theory (MRPT), and coupled-cluster theory (MRCC) [3]. Although

[1] M. L. Abrams and C. D. Sherrill, J. Phys. Chem. A 107, 5611-5616 (2003).

the idea of such methods is rather old,[90–92] significant new approximations and algorithms [3, 5, 6, 82, 93–99] have been developed. While these methods remain too computationally expensive to use on molecules with more than a few heavy atoms, they are nonetheless the only methods capable of accurately describing many chemical processes, particularly bond-breaking and bond-forming reactions.

Multi-reference configuration interaction has been the standard model for determining accurate potential energy surfaces of polyatomic molecules for the last thirty years. Problems with size-consistency still remain, but thus far alternative size-consistent methods are not yet in common use for generating potential energy surfaces of spectroscopic accuracy. Multi-reference perturbation theory has been applied to a number of chemical problems including molecular structure, electronic spectra, and transition metal chemistry. Perhaps the most popular variation of MRPT is the complete-active-space second order perturbation theory (CASPT2) method of Andersson and Roos [93]. Errors in geometries, binding energies, and excitation energies have been systematically studied [100]; however, the error in CASPT2 along a full potential energy curve has not been thoroughly evaluated.

The most straightforward way to determine the error of a given correlation model is to compare it to the exact solution of the electronic Schrödinger equation for the given one-electron basis, which is the full configuration interaction (FCI) result. A series of studies by Bauschlicher in the late 1980s provided FCI energies at a few geometries along the potential energy curves of several small molecules [101–105]. Advances in CI algorithms and computer hardware have made it possible to obtain more complete FCI potential energy curves [74–77, 106, 107] for some simple systems. For example, Olsen and co-workers studied polarized double-zeta FCI potential energy curves for bond breaking in several electronic states of the N_2 molecule [76, 108] and the symmetric dissociation (breaking both bonds) of H_2O. The benchmark FCI results were compared to perturbation theory and coupled-cluster models to indicate how these approximate methods perform for very challenging cases. Such benchmarks are essential for the calibration of new theoretical models meant to describe bond-breaking processes [109–114].

In the present study, we compare to FCI potential energy curves for three molecules

(BH, HF, CH_4) in which a bond to hydrogen is broken. This should represent a common, chemically important process which one might expect to be the easiest type of bond-breaking reaction for standard quantum chemical methods to describe accurately. However, we have recently shown [77] that single-reference methods, even when based on an unrestricted Hartree-Fock reference, are not very accurate for these simple systems; unrestricted coupled-cluster with perturbative triple excitations [UCCSD(T)] yields non-parallelity errors of about 4 kcal/mol. It is therefore of interest to compare the performance of multi-reference approaches for these molecules.

Here we assess the popular complete-active-space self-consistent field (CASSCF) [115, 116] method and complete-active-space second-order perturbation theory (CASPT2) with the nondiagonal zero-order operator [93]. We also consider two multi-reference configuration interaction singles and doubles (MRCISD) methods. One of these, the second-order CI (SOCI) [90], generates all possible singly and doubly substituted configurations from every active space configuration which can be formed by distributing the active electrons among the active orbitals. This is perhaps the most complete type of a MRCISD wave function. The second MRCISD approach considered here is the CISD[TQ] wave function of Schaefer and co-workers [117, 118], which generates all single and double substitutions from the reference set of all singly and doubly substituted configurations which can be formed in the active space. As such, CISD[TQ] may be thought of as an approximation to SOCI in which all configurations which are more than quadruply substituted (relative to the Hartree-Fock reference) are discarded.

Previous high-quality benchmarks for the molecules considered in this study include a cc-pVTZ FCI potential energy curve for BH [119], a cc-pVDZ FCI potential energy curve for HF [119], and 6-311++G(df,p) multi-reference CI results [120] for the breaking of a single C–H bond in CH_4. We have previously assessed [77] several single-reference correlated methods by comparison to FCI for BH, HF, and CH_4 in the aug-cc-pVQZ, 6-31G**, and 6-31G* basis sets, respectively, and we will refer to the FCI results of that study in the present work. Because of additional difficulties in converging CASSCF wave functions with basis sets containing many diffuse functions, for BH we will compare to cc-pVQZ FCI results,

which were also generated in our previous study.

4.3 Theoretical Approach

Here we use the cc-pVQZ, 6-31G**, and 6-31G* basis sets for BH, HF, and CH_4, respectively [34]. We have examined CASSCF [115, 116], CISD[TQ] [118], SOCI [90], and CASPT2 with the nondiagonal zero-order operator [93]. The CISD[TQ] and SOCI computations were performed using CASSCF orbitals.

Two orbital active spaces were used in this study. The first active space is a valence active space, by which we mean that there is one active space orbital of a given irreducible representation for each molecular orbital of the same irreducible representation which can be formed from the valence orbitals on the atoms in the molecule. The valence orbitals of the atoms are of course the usual 1s or hydrogen and 2s, $2p_x$, $2p_y$, $2p_z$ for B-F. The valence active space is commonly used in CASSCF computations, where it is sometimes referred to as full-valence CASSCF. This approach was also introduced by Ruedenberg under the name full optimized reaction space (FORS) [116].

The second active space, which we will refer to as one-to-one or 1:1, includes the occupied valence orbitals plus an active virtual orbital (of the same symmetry) for every occupied valence orbital. This is the active spaced used in generalized valence bond perfect pairing computations, and it is being considered in recent work by Head-Gordon and co-workers [121, 122]. Its use in CASSCF computations dates back at least to a 1980 paper by Roos which shows much better agreement with experiment than full-valence CASSCF for equilibrium properties of the water molecule [123]. The benefit of these two choices of active spaces, full-valence and one-to-one, is that they are *a priori* definitions which can be used to define the active space of a molecule in the absence of any preliminary computations or arbitrary choices of thresholds. These two active spaces were previously compared for the more challenging case of double dissociation in H_2O by Olsen and co-workers [75].

More specifically, the active spaces for the molecules in the study are: BH val=(4e-/3011) and 1:1=(4e-/4000); HF val=(8e-/3011) and 1:1=(8e-/4022); and CH_4 val=1:1=(8e-/62), where the notation indicates (number of active electrons/number of active orbitals per irrep

of the largest Abelian subgroup). For CH_4, definitions the valence and 1:1 active spaces are equivalent. Note that the valence active space is larger than the 1:1 active space for the BH molecule, but it is smaller for HF. The core 1s orbitals were constrained to remain doubly occupied in all cases.

The CASPT2 calculations were performed with MOLCAS 5.2 [124]. All other calculations were performed with the DETCAS and DETCI [82] modules of PSI 3.2 [125]. The potential energy curve for methane was obtained by constraining three C–H bonds to their equilibrium bond length (1.086Å)[126] and the H–C–H angles at the tetrahedral value while stretching a single C–H bond. FCI results are taken from our previous work [77].

4.4 Results and Discussion

The potential energy curves for BH are shown in Figure 1. From the figure, it is clear that all of the methods considered here provide qualitatively correct potential energy curves. This contrasts to the behavior of many single-reference correlation methods, evaluated previously [77]. Although the CASSCF potential energy curves are significantly higher in energy than the exact FCI curve (because of the limited treatment of electron correlation), they have approximately the correct shape. The CASPT2 curve with the smaller active space is also significantly higher than the FCI curve, but again it has the correct shape. The other methods considered are accurate enough that they are hard to distinguish from the FCI curve at this scale. Since the qualitative behavior of the potential energy curves is similar for the other two molecules, we omit figures of the curves for HF and CH_4. More instructive are plots of the errors versus FCI as a function of bond length, which are presented in Figures 2, 3, and 4 for BH, HF, and CH_4, respectively. These are discussed below.

Tables 11, 12, and 13 contain the error versus FCI, and Tables 14, 15, and 16 contain the maximum, minimum, and non-parallelity errors (NPE) for BH, HF, and CH_4, respectively. The NPEs are computed as the difference between the maximum and minimum errors along the potential energy curve, and they provide a measure of how well each method mimics the overall shape of the exact, FCI potential energy curve. An NPE of zero would imply a complete match to the shape of the FCI curve.

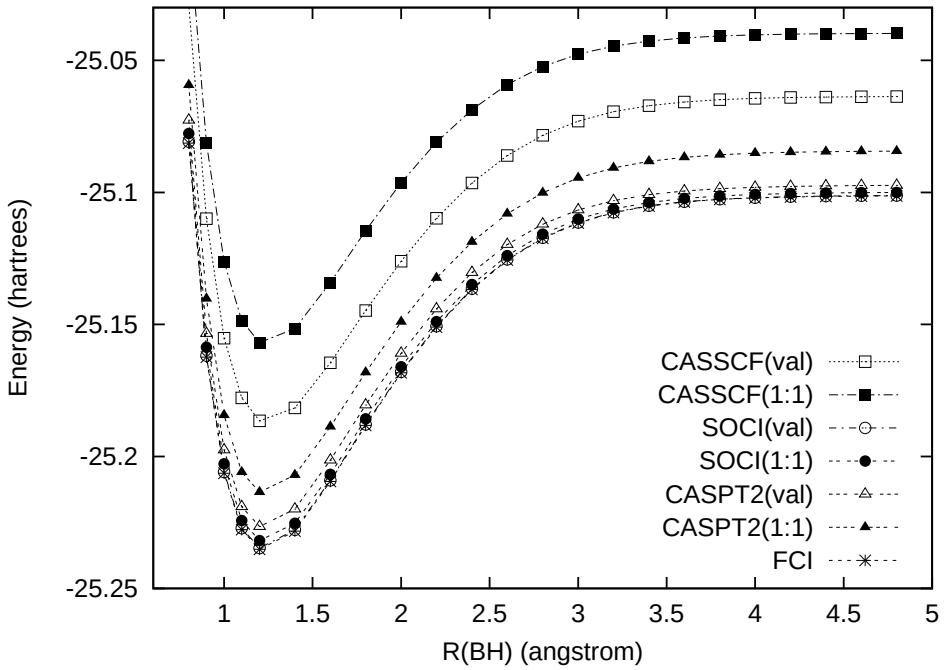

Figure 1: Potential energy curves (in a.u.) for BH using the cc-pVQZ basis set.

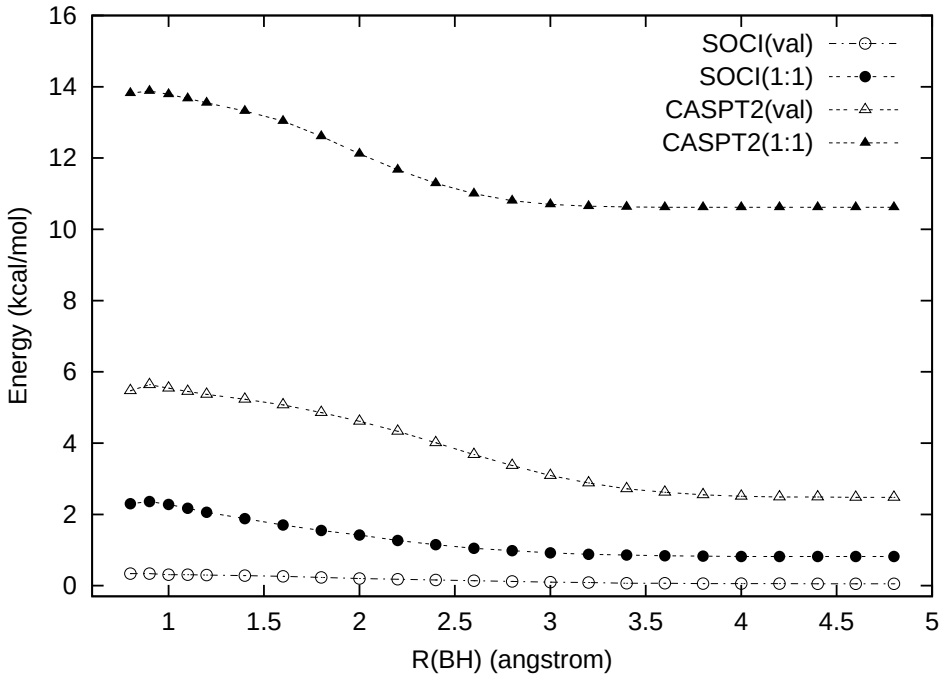

Figure 2: Error versus FCI (in kcal/mol) for BH using the cc-pVQZ basis set. The cc-pVQZ FCI equilibrium bond distance is 1.2260 Å.

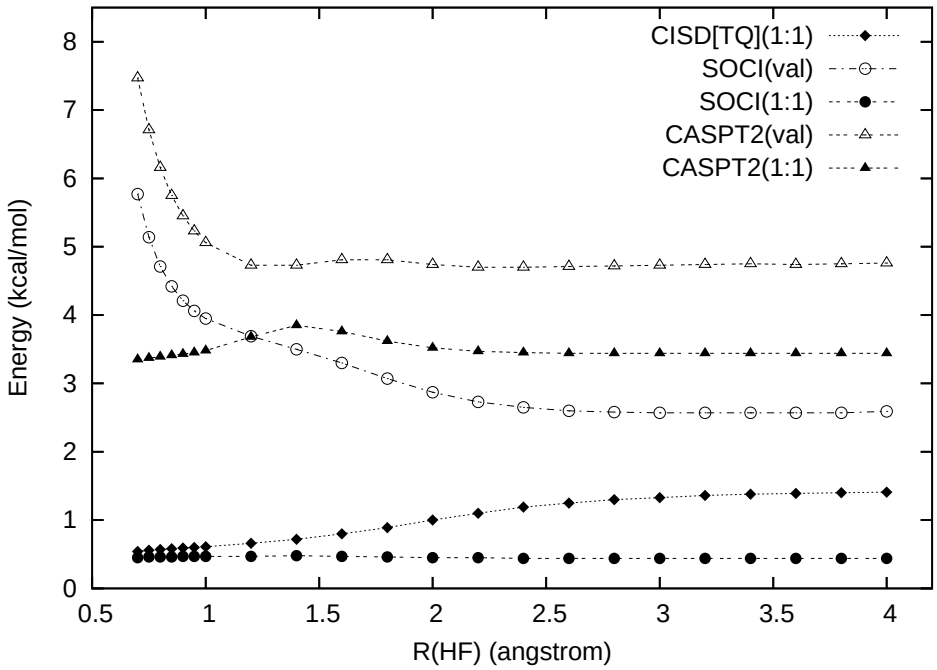

Figure 3: Error versus FCI (in kcal/mol) for HF using the 6-31G** basis set. The 6-31G** FCI equilibrium bond distance is 0.9214 Å.

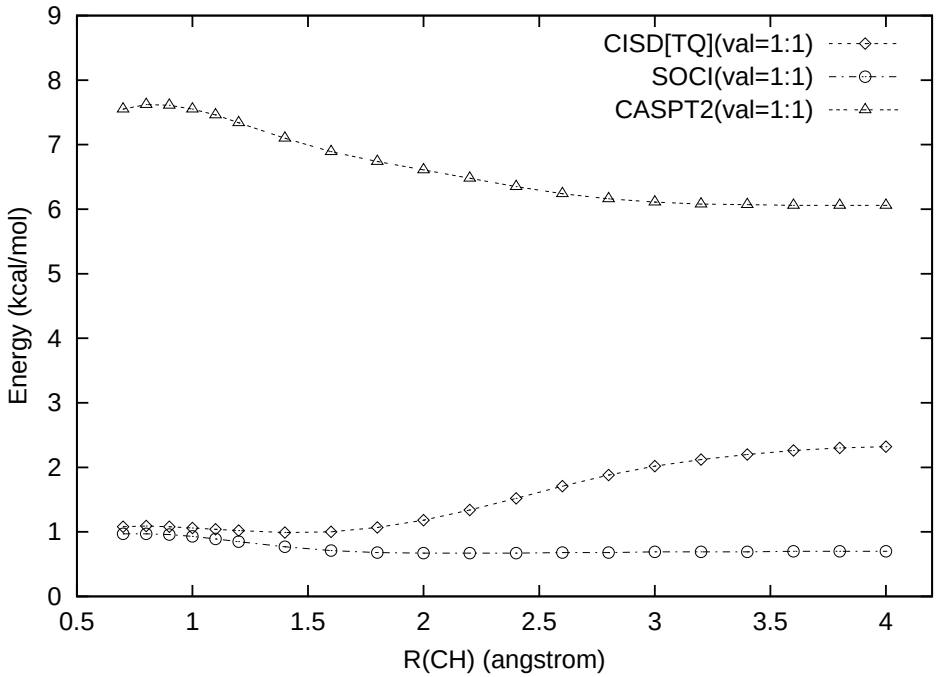

Figure 4: Error versus FCI (kcal/mol) for CH_4 using the 6-31G* basis set. The equilibrium bond distance is 1.086 Å from Ref. [126].

Table 11: FCI energy and error versus FCI (in a.u.) for BH using the cc-pVQZ basis set.

R_{BH}/Å	FCI	CASSCF(val)	CASSCF(1:1)	SOCI(val)	SOCI(1:1)	CASPT2(val)	CASPT2(1:1)
0.80	-25.081355	0.052543	0.081702	0.000543	0.003673	0.008719	0.022021
0.90	-25.162408	0.052480	0.081161	0.000536	0.003764	0.008986	0.022119
1.00	-25.206390	0.051187	0.080188	0.000496	0.003628	0.008822	0.021979
1.10	-25.227709	0.049855	0.079181	0.000490	0.003451	0.008685	0.021784
1.20	-25.235155	0.048658	0.078273	0.000479	0.003286	0.008565	0.021599
1.40	-25.228283	0.046621	0.076734	0.000447	0.002988	0.008338	0.021225
1.60	-25.209491	0.044943	0.075332	0.000408	0.002717	0.008073	0.020788
1.80	-25.188268	0.043511	0.073798	0.000365	0.002477	0.007742	0.020088
2.00	-25.168318	0.042268	0.071984	0.000325	0.002256	0.007346	0.019317
2.20	-25.151002	0.041200	0.070005	0.000289	0.002032	0.006895	0.018594
2.40	-25.136762	0.040307	0.068116	0.000256	0.001833	0.006396	0.017990
2.60	-25.125628	0.039581	0.066450	0.000224	0.001676	0.005871	0.017526
2.80	-25.117376	0.039010	0.065052	0.000193	0.001557	0.005366	0.017217
3.00	-25.111577	0.038575	0.063942	0.000164	0.001471	0.004929	0.017045
3.20	-25.107695	0.038256	0.063116	0.000138	0.001410	0.004589	0.016965
3.40	-25.105194	0.038032	0.062538	0.000119	0.001369	0.004343	0.016933
3.60	-25.103629	0.037878	0.062151	0.000106	0.001342	0.004174	0.016923
3.80	-25.102666	0.037775	0.061901	0.000097	0.001325	0.004065	0.016921
4.00	-25.102081	0.037706	0.061740	0.000091	0.001315	0.004002	0.016921
4.20	-25.101726	0.037660	0.061638	0.000088	0.001308	0.003975	0.016922
4.40	-25.101512	0.037629	0.061573	0.000087	0.001304	0.003962	0.016922
4.60	-25.101383	0.037609	0.061531	0.000086	0.001302	0.003956	0.016922
4.80	-25.101305	0.037594	0.061503	0.000086	0.001300	0.003957	0.016922

Table 12: Error versus FCI (in a.u.) for HF using the 6-31G** basis set.

R_{HF}/Å	CASSCF(val)	CASSCF(1:1)	CISD[TQ](1:1)	SOCI(val)	SOCI(1:1)	CASPT2(val)	CASPT2(1:1)
0.70	0.162876	0.063955	0.000867	0.009195	0.000716	0.011912	0.005343
0.75	0.163972	0.064196	0.000886	0.008197	0.000726	0.010698	0.005374
0.80	0.164811	0.064369	0.000904	0.007510	0.000733	0.009819	0.005404
0.85	0.165401	0.064491	0.000920	0.007037	0.000738	0.009168	0.005434
0.90	0.165754	0.064589	0.000936	0.006706	0.000742	0.008688	0.005466
0.95	0.165890	0.064691	0.000951	0.006469	0.000744	0.008328	0.005504
1.00	0.165824	0.064822	0.000967	0.006293	0.000746	0.008056	0.005552
1.20	0.163754	0.065908	0.001045	0.005878	0.000756	0.007540	0.005872
1.40	0.159260	0.067861	0.001147	0.005585	0.000761	0.007543	0.006140
1.60	0.153437	0.069764	0.001267	0.005257	0.000747	0.007666	0.005996
1.80	0.147734	0.070855	0.001421	0.004894	0.000731	0.007659	0.005766
2.00	0.143307	0.071317	0.001593	0.004572	0.000719	0.007561	0.005612
2.20	0.140478	0.071503	0.001756	0.004347	0.000712	0.007494	0.005529
2.40	0.138900	0.071590	0.001891	0.004216	0.000709	0.007484	0.005494
2.60	0.138081	0.071633	0.001993	0.004146	0.000707	0.007501	0.005483
2.80	0.137669	0.071651	0.002069	0.004112	0.000706	0.007521	0.005481
3.00	0.137462	0.071655	0.002125	0.004095	0.000706	0.007537	0.005484
3.20	0.137360	0.071654	0.002167	0.004089	0.000706	0.007551	0.005485
3.40	0.137309	0.071651	0.002197	0.004088	0.000705	0.007565	0.005486
3.60	0.137285	0.071648	0.002218	0.004090	0.000705	0.007556	0.005486
3.80	0.137273	0.071646	0.002232	0.004094	0.000705	0.007576	0.005486
4.00	0.137268	0.071645	0.002242	0.004133	0.000705	0.007579	0.005486

Table 13: Error versus FCI (in a.u.) for CH_4 using the 6-31G* basis set.

R_{CH}/Å	CASSCF(val=1:1)	CISD[TQ](val=1:1)	SOCI(val=1:1)	CASPT2(val=1:1)
0.70	0.082722	0.001716	0.001550	0.012026
0.80	0.081922	0.001733	0.001551	0.012138
0.90	0.080894	0.001722	0.001523	0.012128
1.00	0.079746	0.001696	0.001475	0.012034
1.10	0.078547	0.001663	0.001415	0.011883
1.20	0.077354	0.001628	0.001350	0.011698
1.40	0.075191	0.001583	0.001226	0.011309
1.60	0.073614	0.001601	0.001134	0.010986
1.80	0.072781	0.001702	0.001082	0.010742
2.00	0.072627	0.001885	0.001061	0.010533
2.20	0.072942	0.002137	0.001060	0.010326
2.40	0.073469	0.002429	0.001069	0.010122
2.60	0.073997	0.002724	0.001081	0.009943
2.80	0.074414	0.002991	0.001091	0.009815
3.00	0.074697	0.003213	0.001099	0.009734
3.20	0.074872	0.003384	0.001104	0.009689
3.40	0.074971	0.003510	0.001107	0.009667
3.60	0.075024	0.003599	0.001109	0.009657
3.80	0.075051	0.003659	0.001110	0.009652
4.00	0.075064	0.003699	0.001110	0.009650

It is well known that the CASSCF wave function provides qualitatively correct potential energy surfaces if a proper active space is chosen, because it describes the interaction between important near-degenerate configurations, called non-dynamical correlation. However, the method does not contain enough configurations to accurately describe the usual short-range electron-electron repulsions, called dynamical correlation. Since more electrons are closer together near equilibrium than at the dissociation limit, the degree of dynamical correlation is larger there, and so are the CASSCF errors. This is seen quantitatively in Tables 11-13. The only exception is for HF with the 1:1 active space, where the CASSCF error is fairly constant and rises slightly with distance. Here, the 1:1 active space is sufficient to capture a significant portion of the dynamical correlation near equilibrium.

Given that the degree of dynamical correlation is larger near equilibrium, one would expect that increasing the size of the active space would cause a greater improvement in the error at equilibrium than at dissociation. The non-parallelity errors in Tables 14 and 15

Table 14: Maximum, minimum, and non-parallelity error (in kcal/mol) for BH using the cc-pVQZ basis set. Values in parentheses indicate the corresponding bond distance (in Å).

Method	Max error	Min error	NPE
CASSCF(val)	32.97 (0.80)	23.59 (4.80)	9.38
CASSCF(1:1)	51.27 (0.80)	38.59 (4.80)	12.68
SOCI(val)	0.34 (0.80)	0.05 (4.40)	0.29
SOCI(1:1)	2.36 (0.90)	0.82 (4.00)	1.54
CASPT2(val)	5.64 (0.90)	2.48 (4.60)	3.16
CASPT2(1:1)	13.88 (0.90)	10.62 (3.60)	3.26

Table 15: Maximum, minimum, and non-parallelity error (in kcal/mol) for HF using the 6-31G** basis set. Values in parentheses indicate the corresponding bond distance (in Å).

Method	Max error	Min error	NPE
CASSCF(val)	104.10 (0.95)	86.14 (3.80)	17.96
CASSCF(1:1)	44.96 (2.80)	40.13 (0.70)	4.83
SOCI(val)	5.77 (0.70)	2.57 (3.00)	3.20
SOCI(1:1)	0.48 (1.40)	0.44 (2.40)	0.04
CISD[TQ](1:1)	1.41 (4.00)	0.54 (0.70)	0.87
CASPT2(val)	7.47 (0.70)	4.70 (2.20)	2.77
CASPT2(1:1)	3.85 (1.40)	3.35 (0.70)	0.50

Table 16: Maximum, minimum, and non-parallelity error (in kcal/mol) for CH_4 using the 6-31G* basis set. Values in parentheses indicate the corresponding bond distance (in Å).

Method	Max error	Min error	NPE
CASSCF(val=1:1)	51.91 (0.7)	45.57 (2.0)	6.34
CISD[TQ](val=1:1)	2.32 (4.0)	0.99 (1.4)	1.33
SOCI(val=1:1)	0.97 (0.7)	0.67 (2.0)	0.30
CASPT2(val=1:1)	7.62 (0.8)	6.06 (3.6)	1.56

verify this expectation, keeping in mind that the valence active space is larger for BH but smaller for HF. Enlarging the active space from 1:1 to valence for BH decreases the error at 1.2 Å by 18.6 kcal/mol and at 4.8 Å by 15.0 kcal/mol. Likewise, for HF enlarging the active space from valence to 1:1 reduces the error at 0.9 Å by 63.5 kcal/mol and at 4.0 Å by 41.2 kcal/mol.

Across all three molecules, the CASSCF non-parallelity error is as small as 4.8 kcal/mol and as large as 18.0 kcal/mol. This is roughly comparable to the NPEs for CCSD based on

RHF orbitals (8-13 kcal/mol) and represents a definite improvement over UHF or UMP2, but not UCCSD, UCCSD(T), or UB3LYP [77]. However, for more challenging bond breaking cases, CASSCF should continue to work as well, while the quality of the single-reference approaches will degrade, as demonstrated by Olsen and co-workers [75] for the simultaneous breaking of both bonds in H_2O. We plan to examine bond breaking in additional molecules in future work.

CISD[TQ] is a restricted active space [57] variant of MRCI where RAS I is composed of active occupied orbitals and RAS II is composed of the active unoccupied orbitals. The inactive virtual orbitals comprise RAS III. The configurations are selected according to the following criteria: a maximum of two electrons are allowed in RAS III, and a maximum of 4 holes are allowed in RAS I. There are no restrictions on the occupancy of the orbitals in RAS II. These rules are equivalent to generating a MRCISD in which the reference configurations are all singles and doubles within the active space. Results are not given for CISD[TQ] for BH because for 4 valence electrons, CISD[TQ] is equivalent to SOCI (discussed below). Likewise, for the small valence active space in HF (only one active virtual orbital), again CISD[TQ] and SOCI are equivalent.

It has been shown [118] that the error versus FCI for CISD[TQ] along a potential energy curve is the opposite of CASSCF: the maximum error occurs near dissociation while the minimum error occurs near equilibrium. Figures 3 and 4 verify this trend. The error at 0.7 Å is quite small for HF and CH_4 (0.5 and 1.1 kcal/mol, respectively). The error begins to increase at 1.0 Å for HF, but decreases slightly for CH_4 before beginning to climb at 1.8 Å. The non-parallelity error does achieve chemical accuracy (1.0 kcal/mol) for HF. In both cases, the NPE is two to four times less than UCCSD(T) [77].

The SOCI wave function can also be defined in the restricted active space framework by placing the active space orbitals in RAS II and the inactive virtual orbitals in RAS III. It is not necessary to employ the RAS I orbital space in this case. Configurations are selected by allowing all possible distributions of electrons in RAS II but a maximum of two electrons in RAS III. Because this is equivalent to a MRCISD in which every possible active space configuration is used as a reference, SOCI represents an idealized limit of MRCISD which

would be very difficult to employ for all but the smallest chemical systems.

When the active space is small, SOCI describes the region near dissociation more accurately than that around equilibrium. This trend is seen for BH with SOCI (1:1) and HF with SOCI (val); the associated non-parallelity errors (1.5 and 3.2 kcal/mol) are surprisingly large considering the very extensive description of electron correlation in SOCI. However, it must be kept in mind that the valence active space contains only one active unoccupied orbital for HF, and no b_1 or b_2 orbitals for BH, which are important for dynamical correlation near equilibrium.

When one uses the larger active space in each case, the SOCI parallels the FCI limit very well, as seen for BH with SOCI (val), HF with SOCI (1:1), and CH_4 with SOCI (val=1:1). The non-parallelity errors for SOCI with the larger active space (0.29-0.04 kcal/mol) are about an order of magnitude smaller than for UCCSD(T) for these molecules [77]. The significantly larger non-parallelity errors for the smaller active space show that even for very extensive MRCISD computations, the active space must be chosen carefully if sub-chemical accuracy is to be achieved.

Since its inception in 1990 [127], many papers have been published using the CASPT2 variant of MRPT. Since CASPT2 is approximately size-consistent [128] and will generally be more computationally efficient than MRCI, it may be attractive as an alternative approach for the determination of accurate potential energy surfaces. Previous studies [127, 129] have shown that the CASPT2 error mimics the trend of SOCI–maximum error near equilibrium and minimum error near dissociation. This trend is seen here in Figures 2 and 3. As one would expect, the error for CASPT2 decreases with the larger active space. In the case of HF, the larger active space also leads to a smaller NPE, but in BH the NPE is hardly improved. It is interesting to note that the NPEs for CASPT2 are less sensitive to the active space than are the NPEs for SOCI. Except for HF with the valence active space, the CASPT2 NPEs are always larger than the SOCI NPEs, by a factor of around 2-10. This indicates that SOCI or similar large MRCISD wave functions are preferable to CASPT2 for obtaining very accurate potential energy surfaces, such as might be required for high-accuracy vibrational level prediction. Nevertheless, the CASPT2 errors are modest,

and they tend to compare well with CISD[TQ]. Compared to the single-reference results of our previous study, we find that CASPT2 does as well as UCCSD(T) for BH and much better for HF and CH_4. We expect the improvement over UCCSD(T) to be larger for more challenging cases such as the homolytic dissociation of bonds between two non-hydrogen atoms.

4.5 Conclusions

We have assessed various multi-reference methods for their performance in bond breaking reactions in BH, CH_4, and HF. This represents the first detailed evaluation of multi-reference methods across the entire potential energy curve for single bond breaking reactions, and it is made possible by our determination of computationally demanding full configuration interaction potential curves, which represent the exact solution for the given basis set. Although breaking bonds to hydrogen should represent one of the easiest types of bond breaking reactions for theoretical methods, nevertheless our recent evaluation of single-reference methods indicated surprisingly large errors [77]. For the present molecules, the multi-reference methods including dynamical correlation (CISD[TQ], SOCI, and CASPT2) all give more accurate results than the best standard single-reference method, UCCSD(T), even for these simple reactions.

With current technology, multi-reference configuration interaction is the only method in common use for determining spectroscopic quality potential energy surfaces. As the size of the system increases, the quality of the MRCI wave function will degrade, leading to a need for size-consistent (or at least approximately size-consistent) methods, including multi-reference perturbation theory or multi-reference coupled-cluster theory. The CASPT2 approach is a widely used multi-reference perturbation theory method which, according to the present study, appears to be somewhat less sensitive to the choice of active space than MRCI. However, the perturbative treatment of dynamical electron correlation does not seem as effective as extensive MRCI for the small systems considered here. Unfortunately, the prospects of improving the CASPT2 model by employing higher orders of perturbation theory does not seem promising, as Olsen and co-workers [130] have shown that the

multi-reference perturbation expansion does not converge. Nevertheless, if the proper active orbital space is chosen, CASPT2 is still preferred to spin-unrestricted single-reference methods for describing potential energy curves.

Comparing active space definitions, valence or one-to-one, demonstrates that the best definition for a given molecule is that which produces the larger active space. The smaller active space led to relatively large errors in some cases. The ideal active space for multi-reference computations—which may be too large to employ in practice—would be one which is the union of the valence and one-to-one active spaces: i.e., it should include all valence orbitals *and* ensure the presence of at least one active unoccupied orbital per active occupied orbital.

CHAPTER 5

FULL CONFIGURATION INTERACTION BENCHMARKS: III

5.1 Abstract

The C_2 molecule exhibits unusual bonding and several low-lying excited electronic states, making the prediction of its potential energy curves a challenging test for quantum chemical methods. We report full configuration interaction results for the X $^1\Sigma_g^+$, B $^1\Delta_g$, and B$'$ $^1\Sigma_g^+$ states of C_2 which exactly solve the electronic Schrödinger equation within the space spanned by a 6-31G* basis set. Within the D_{2h} subgroup used by most electronic structure programs, these states all have the same symmetry (^1A_g), and all three states become energetically close for interatomic distances beyond 1.5 Å. The quality of several single-reference *ab initio* methods is assessed by comparison to the benchmark results. Unfortunately, even coupled-cluster theory through perturbative triples using an unrestricted Hartree-Fock reference [UCCSD(T)] exhibits large non-parallelity errors (>20 kcal/mol) for the ground state. The excited states are not accurately modeled by any commonly-used single-reference method, nor by configuration interaction including full quadruple substitutions. The present benchmarks will be helpful in assessing theoretical methods designed to break bonds in ground and excited electronic states.[1]

5.2 Introduction

The C_2 molecule is a common intermediate in combustion reactions and is central to the chemistry of the interstellar medium. As such, its spectroscopy has been heavily studied. The X $^1\Sigma_g^+$ ground electronic state exhibits very unusual bonding, having two π bonds but no σ bond. There is also an exceptionally low-lying electronic excited state, the $^3\Pi_u$ state,

[1] M. L. Abrams and C. D. Sherrill, J. Chem. Phys. 121, 9211-9219 (2004).

only 716 cm^{-1} higher in energy, and about 16 other electronic states have been observed [131]. The molecule's unusual bonding and the presence of many low-lying electronic excited states make it a challenging target for theoretical studies.

In the present article, C_2 is used to test the reliability of quantum chemical methods in bond-breaking reactions. Unfortunately, even sophisticated *ab initio* methods can have severe difficulties accurately modeling bond-breaking or bond-forming processes, because the single Slater determinant upon which they are usually built is incapable of properly describing degeneracies among electron configurations which arise at the dissociation limit. It is well known that a restricted Hartree-Fock (RHF) determinant includes unphysical, ionic terms at dissociation for homolytic cleavage of covalent bonds, artificially increasing the energy at large distances. Allowing different orbitals for different spins in the unrestricted Hartree-Fock (UHF) method can yield a potential energy curve which is qualitatively correct, but which is often quantitatively poor; moreover, the wave function is no longer an eigenfunction of the spin operator $\hat{S}^2$. The deficiencies in RHF or UHF wave functions are so severe that, in general, they cannot be adequately corrected by the addition of electron correlation via standard approaches such as single-reference perturbation theory or coupled-cluster theory. In principle, multi-reference methods such as complete-active-space self-consistent-field with second-order perturbation theory corrections (CASPT2) [93] can accurately model such problems; unfortunately, however, they are more computationally expensive, harder to use by non-experts, and not as widely available in program packages. Thus it is critical to assess the reliability of both single- and multi-reference methods for bond-breaking reactions.

The most straightforward way to determine the accuracy of predicted potential energy curves is to compare them to exact results for a given one-electron basis set. This is possible using full configuration interaction (FCI), which solves the electronic Schrödinger equation exactly by including all possible Slater determinants of the appropriate symmetry which can be formed from the given basis set. Because the number of interacting determinants grows factorially with the number of electrons or basis functions, unfortunately FCI computations are usually only feasible for small molecules with modest basis sets. Several studies in

the 1980s by Bauschlicher and co-workers presented FCI energies at a small number of geometries along a potential energy curve [101–105]. In the last decade, more complete FCI potential curves have been presented for a few systems [74–77, 106–108], and benchmarks such as these have been useful in testing new theoretical methods designed for bond-breaking reactions [109–114, 132].

Among the FCI studies of potential energy curves, two notable ones by Olsen and co-workers examine several states of N_2 [76, 108] and the symmetric dissociation of H_2O.[75] By breaking two bonds (H_2O) or three bonds (N_2) simultaneously, these potential energy curves represent particularly challenging test cases for standard quantum chemical methods. Surprisingly, very little FCI data was available on less challenging cases, such as breaking single bonds to hydrogen, until recently. In a study of bond breaking in BH, HF, and CH_4, which should present the least challenge to single-reference theories, unexpectedly large errors were found for otherwise very sophisticated methods [77]. For example, the non-parallelity errors (NPEs), determined as the difference between the maximum and minimum errors along the curve, for UHF-based coupled-cluster theory with singles, doubles, and perturbative triple excitations [UCCSD(T)][133] were as large as 4 kcal/mol. The CASPT2 approach, although it uses a much simpler model of dynamical electron correlation, improved non-parallelity errors for these reactions to about 3 kcal/mol when a large active space was used [134].

In this study, we provide FCI benchmarks for the ground state of C_2, which having two bonds should be a challenging test case, but not as challenging as N_2. We also help alleviate the extreme scarcity of FCI curves for excited states [76, 107, 108] by presenting results for the B $^1\Delta_g$ and B$'$ $^1\Sigma_g^+$ states, which both dissociate to the same limit, 2 C (^3P). The B $^1\Delta_g$ state is particularly interesting because, in the D_{2h} computational subgroup used by most programs, it has the same symmetry (^1A_g) as the ground state. Hence, quantum chemistry programs may happen to find a solution representing the B $^1\Delta_g$ state rather than the X $^1\Sigma_g^+$ ground state. As we will show, the FCI energy of the the B $^1\Delta_g$ state actually drops below that of the the X $^1\Sigma_g^+$ state at larger distances. The reliability of various standard single-reference methods is assessed by comparison to the exact results for the ground state. In future work, we will compare results from more elaborate, multi-reference methods which

are specifically designed to describe bond-breaking reactions.

Previous, high-quality computations of the twelve lowest singlet and triplet electronic states of C_2 have been obtained by Halvick and co-workers [135] using contracted multi-reference configuration interaction wave functions and a large cc-pV5Z basis set. Although these results should be quite reliable, they nevertheless fall short of an exact treatment of electron correlation, and hence they are not ideal as benchmarks for electronic structure methods. Full CI computations using a $3s2p1d$ atomic natural orbital basis were reported for one geometry each of the X $^1\Sigma_g^+$, a $^3\Pi_u$, and b $^3\Sigma_g^-$ states by Bauschlicher and co-workers [136], who also reported spectroscopic constants of these states using approximate methods and larger basis sets. Christiansen and co-workers [137] reported FCI vertical excitation energies for four singlet states of C_2 in a cc-pVDZ basis augmented by diffuse s and p functions in an examination of the quality of coupled-cluster excited state methods. Spectroscopic constants for the X $^1\Sigma_g^+$ state, including the vibrational anharmonicity $\omega_e x_e$ and the vibration-rotation coupling constant α_e, have also been reported using FCI and various polarized double-zeta basis sets [73, 138]. FCI potential curves obtained with a DZ basis set (without polarization) have been reported by Piecuch and co-workers and used for benchmarking the method of moments and completely-renormalized coupled-cluster methods [139, 140]. Here we present the first report of the FCI potential energy curve for the X $^1\Sigma_g^+$ ground state using a more reliable, polarized double-zeta basis set. We also present the first FCI study of the avoided crossing of the X $^1\Sigma_g^+$ and B$'$ $^1\Sigma_g^+$ states, and the first FCI study of the B $^1\Delta_g$ state potential energy curve. The present results support earlier conclusions that the standard single-reference coupled-cluster methods can fail badly for reactions breaking multiple bonds [139–142].

5.3 Theoretical Approach

We have used the standard basis set 6-31G* [79, 143], which is, perhaps surprisingly, among the most reliable polarized double-zeta basis sets for FCI benchmarking [138]. All six Cartesian d-type polarization functions were used. Energies have been evaluated for a large

number of interatomic distances using FCI, second-order Møller-Plesset perturbation theory (MP2), configuration interaction with single and double substitutions (CISD), coupled-cluster theory with singles and doubles (CCSD),[29] and CCSD plus perturbative triples [CCSD(T)].[133] To obtain a UHF energy lower than the RHF energy, it is necessary in this case to break the inversion symmetry and perform the computations in the C_{2v} subgroup (UHF solutions were obtained through Hartree-Fock stability analysis, following the eigenvectors corresponding to negative eigenvalues). The C_2 molecule is unusual in that this procedure yielded lower-energy UHF solutions not only at large internuclear distances, but at all geometries considered (this does not happen, for example, in H_2). For convenience, methods based on unrestricted orbitals are denoted with a prefix 'U', as in UMP2, UCCSD, etc. The frozen core approximation was used in all correlated wave functions. All restricted computations were run using the PSI 3.2 package [144], and all unrestricted computations were run using ACES II [42]. FCI procedures employed the latest version of our DETCI program [82]. For the electronic states considered, the FCI wave function consisted of 52,407,353 determinants.

States were assigned as $^1\Sigma_g^+$ or $^1\Delta_g$ by examination of the leading CI coefficients; the real component of the $^1\Delta_g$ state must have the $|\cdots 1\pi_x^2 3\sigma_g^2\rangle$ and $|\cdots 1\pi_y^2 3\sigma_g^2\rangle$ determinants appear with equal coefficients but different signs (here we use labels σ and π for readability, but strictly speaking the computations are run under the D_{2h} point group and the labels of that group's irreducible representations would apply). For $^1\Sigma_g^+$ states, these determinants appear with equal coefficients and the same sign. For this project, we adapted our code to allow the user to filter out unwanted electronic states by specifying the desired phase between any two determinants. It should also be mentioned that it is possible to obtain higher-multiplicity spin states inadvertently in a determinant-based code. We can easily exclude $M_s = 0$ components of triplets because we use the time reversal symmetry [57] $C(I_\alpha, I_\beta) = (-1)^S C(I_\beta, I_\alpha)$, where I_α and I_β label the α and β strings [145]. However, quintets may appear, because $S = 2$ has the same phase as $S = 0$. We computed $< S^2 >$ values explicitly to ensure that only singlets were considered. This is not normally a problem, but here we observed some low-lying quintet roots for truncated CI wave functions

at large distances.

5.4 Results and Discussion

5.4.1 FCI results

The potential energy curves for the X $^1\Sigma_g^+$, B $^1\Delta_g$, and B′ $^1\Sigma_g^+$ states of C_2 are presented in Figure 5, and the total energies are given in Table 17. The B $^1\Delta_g$ state has a more shallow potential energy well and crosses below the X $^1\Sigma_g^+$ state around 1.7 Å, and the two states remain very close in energy (with differences of less than 5 millihartrees, or 3 kcal/mol) as they approach the dissociation limit. The B′ $^1\Sigma_g^+$ state is below the B $^1\Delta_g$ state at short distances, but they cross around 1.15 Å. Near 1.6 Å, the B′ $^1\Sigma_g^+$ state begins to rise in energy relative to the other two (it cannot cross the X $^1\Sigma_g^+$ state because these two states have the same symmetry). At larger distances, the B′ $^1\Sigma_g^+$ state again approaches the other two states. All three states approach the same asymptotic limit 2 C (^3P) and are nearly degenerate at 2.8 Å. At intermediate distances, the closest approach between the two $^1\Sigma_g^+$ states comes near 1.6 Å, which is the same value obtained from the cc-pV5Z multireference CI data of Halvick and co-workers [135]. This indicates that some of the qualitative features of the FCI potential energy curves are accurately captured even using the modest 6-31G* basis set. Spectroscopic constants for the ground state computed at the 6-31G* FCI level of theory are also in good agreement with experiment [138].

Near equilibrium, the ground state wave function is dominated by the configuration $|(core)2\sigma_g^2 2\sigma_u^2 1\pi_x^2 1\pi_y^2\rangle$, although there is a strong mixing (coefficient of 0.33 at 1.25 Å) with a second configuration, $|(core)2\sigma_g^2 1\pi_x^2 1\pi_y^2 3\sigma_g^2\rangle$. Such a large coefficient for the second most important configuration is quite unusual for the ground state wave function of a molecule near its equilibrium geometry. At this same geometry, the B $^1\Delta_g$ state is dominated by $|(core)2\sigma_g^2 2\sigma_u^2 1\pi_x^2 3\sigma_g^2\rangle$ - $|(core)2\sigma_g^2 2\sigma_u^2 1\pi_y^2 3\sigma_g^2\rangle$, while the B′ $^1\Sigma_g^+$ has these two determinants dominant but with the same sign.

By 1.6 Å, the determinant $|$ (core) 2 σ_g^2 1 π_x^2 1 π_y^2 3 σ_g^2 $\rangle$ becomes much less important to the ground-state wave function, and the determinants $|(core)2\sigma_g^2 2\sigma_u^2 1\pi_x^2 3\sigma_g^2\rangle$ and $|(core)2\sigma_g^2 2\sigma_u^2 1\pi_y^2 3\sigma_g^2\rangle$ are now much more important. The B $^1\Delta_g$ and B′ $^1\Sigma_g^+$ states remain

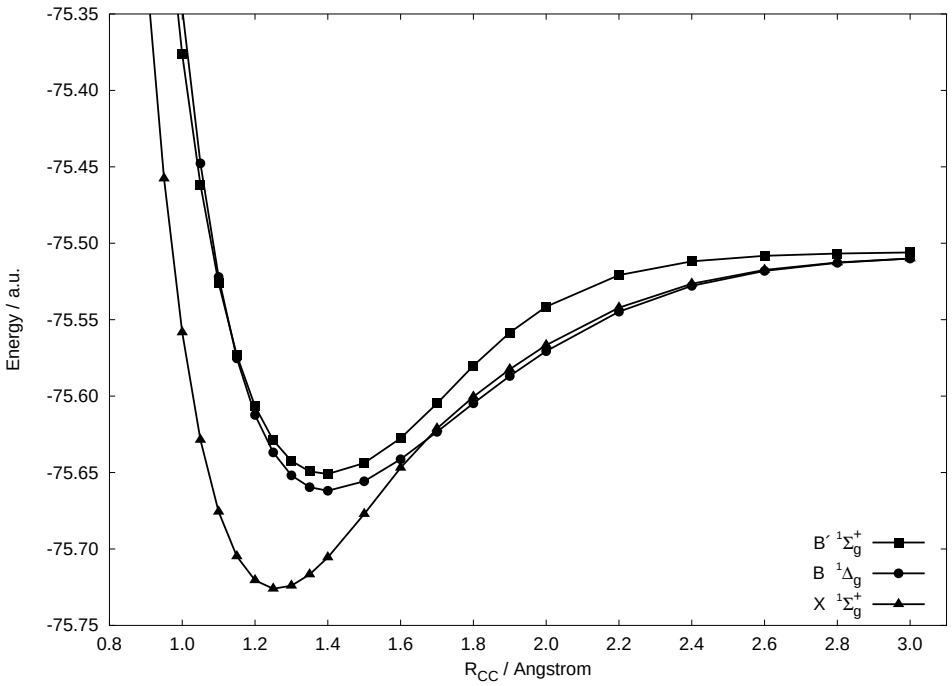

Figure 5: Full configuration interaction potential energy curves using the 6-31G* basis set for the X $^1\Sigma_g^+$, B $^1\Delta_g$, and B′ $^1\Sigma_g^+$ states of C_2.

qualitatively similar to their form near the ground-state equilibrium geometry. However, by 1.80 Å, the B $^1\Delta_g$ state has dropped below the X $^1\Sigma_g^+$ state, and the character of the two $^1\Sigma_g^+$ states is reversed due to an avoided crossing. Note that this avoided crossing leads to a rather unusual shape for the ground state potential energy curve, which does not look like a typical Morse potential. We will observe below that the approximate, single-reference methods have great difficulty reproducing this shape. For all three states, the contribution of additional determinants grows with increasing internuclear separation, and by 2.8 Å, the number of determinants with coefficients greater than 0.20 in each state are: 8 (B $^1\Delta_g$), 6 (X $^1\Sigma_g^+$), and 6 (B′ $^1\Sigma_g^+$).

5.4.2 Standard Single-Reference Results for the X $^1\Sigma_g^+$ State

Correlated wave functions based on a single RHF reference are compared to FCI for the ground state potential energy curve in Figure 6. As one would expect, the behavior of RHF is completely incorrect, with a dissociation energy which is unrealistically high. Adding a description of electron correlation via CISD leads to a curve which is vastly improved but

Table 17: FCI energies (in a.u.) for C_2 using the 6-31G* basis set.

R_{CC}/Å	X $^1\Sigma_g^+$	B $^1\Delta_g$	B′ $^1\Sigma_g^+$
0.90	-75.317618	-75.028726	-75.117717
0.95	-75.457665	-75.210060	-75.264774
1.00	-75.558335	-75.346351	-75.376449
1.05	-75.628645	-75.447724	-75.461663
1.10	-75.675637	-75.522009	-75.526003
1.15	-75.704813	-75.575291	-75.573273
1.20	-75.720475	-75.612329	-75.606636
1.25	-75.725995	-75.636861	-75.628883
1.30	-75.724026	-75.651835	-75.642414
1.35	-75.716657	-75.659574	-75.649224
1.40	-75.705544	-75.661902	-75.650929
1.50	-75.677127	-75.655726	-75.643794
1.60	-75.646930	-75.641311	-75.627561
1.70	-75.621163	-75.623392	-75.604839
1.80	-75.600442	-75.604708	-75.580101
1.90	-75.582417	-75.586820	-75.558438
2.00	-75.566646	-75.570609	-75.541479
2.20	-75.542142	-75.544773	-75.520806
2.40	-75.526459	-75.527876	-75.511848
2.60	-75.517449	-75.518088	-75.508225
2.80	-75.512568	-75.512809	-75.506703
3.00	-75.509925	-75.509987	-75.506025

nevertheless qualitatively incorrect at large distances; we expect a reasonable description of the simultaneous breaking of two bonds to require at least quadruple excitations. The effective, higher-order excitations included in coupled-cluster theory (products of single and double excitations) make the CCSD curve much better than the CISD curve, but even so, there are very large differences between CCSD and FCI, and the difference grows at intermediate C–C distances. Perturbation theory, whether for the doubles in MP2 or for the triples in CCSD(T), fails badly — the correlation energy becomes unphysically large (energies below FCI are possible for perturbation and coupled-cluster theories because they are non-variational). Thus none of these standard, single-reference methods yields a potential curve which is close to the exact solution.

Qualitatively, the behavior of the RHF-based correlated methods for the ground state of C_2 is not so different than that seen for the much easier case of breaking single bonds

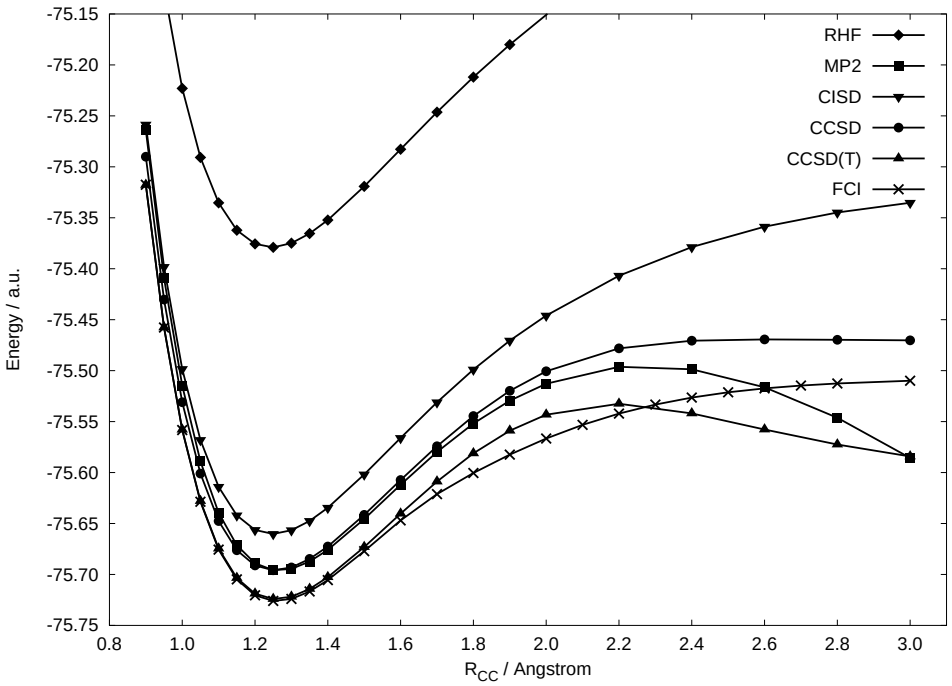

Figure 6: Potential energy curves for X $^1\Sigma_g^+$ C_2 in a 6-31G* basis set using various approximate correlation methods with an RHF reference.

to hydrogen atoms [77]. In both cases, MP2 and CCSD(T) are qualitatively incorrect at large distances; however, for breaking bonds to hydrogen, CCSD(T) levels off below the FCI energy at large distances, whereas it appears to diverge for C_2. In both cases, CCSD is the best of the methods considered here, but it falls too high above the exact curve at dissociation. For the single-bond examples, the error in CCSD leveled off but was greatest at large distances, whereas for C_2 the maximum error is found at intermediate distances.

The errors may be examined more quantitatively in Figure 7, which presents the errors versus FCI as a function of distance. One desires a perfectly flat error curve, which would indicate a mere shifting of the potential. None of the error curves in Figure 7 are close to being flat. The non-parallelity error (NPE), computed as the difference between the maximum and minimum errors found along the curve, is reported for each method in Table 18. Even CCSD, the best of the methods currently considered, exhibits errors which range from 17 kcal/mol near equilibrium to 41 kcal/mol at 2.0 Å, with an NPE of 24 kcal/mol, which is far greater than the 8-13 kcal/mol observed for breaking bonds to the hydrogen

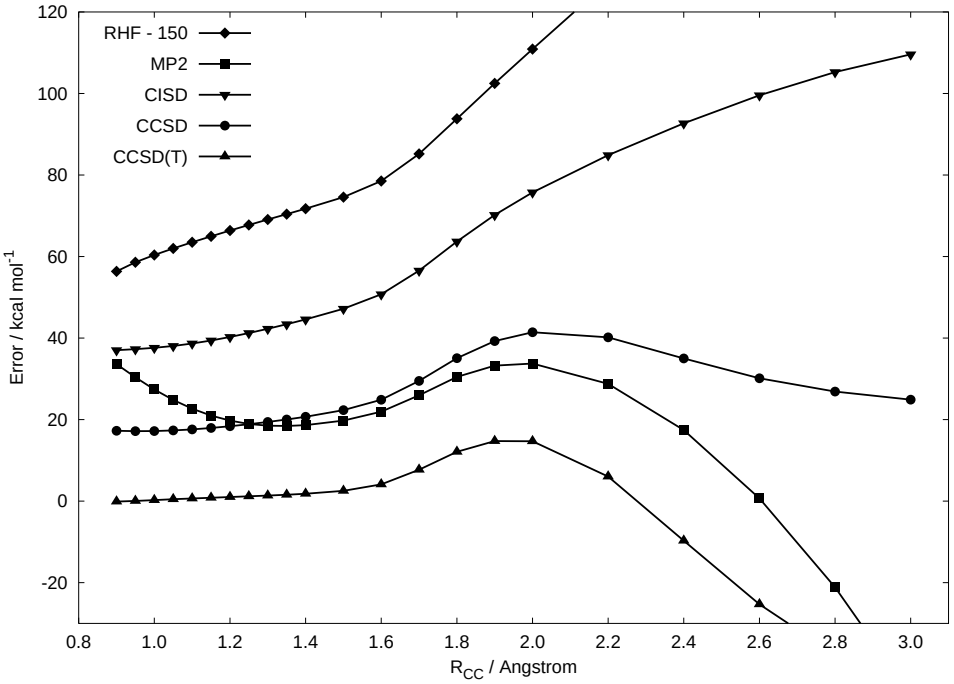

Figure 7: Errors in potential energies for X $^1\Sigma_g^+$ C_2 using various approximate correlation methods with an RHF reference and a 6-31G* basis. RHF errors have been shifted down by 150 kcal/mol.

Table 18: Maximum, minimum, and non-parallelity error (in kcal/mol) versus FCI for the X $^1\Sigma_g^+$ state of C_2 using the 6-31G* basis set. Values in parentheses indicate the corresponding bond distance (in Å).

Method	Max error	Min error	NPE
RHF	339.3 (3.00)	206.4 (0.90)	132.9
MP2	33.8 (2.00)	-47.5 (3.00)	81.3
CISD	109.6 (3.00)	37.0 (0.90)	72.6
CCSD	41.4 (2.00)	17.2 (0.95)	24.3
CCSD(T)	14.7 (1.90)	-46.6 (3.00)	61.3
UHF	142.4 (1.30)	93.7 (3.00)	48.7
UMP2	61.4 (1.80)	20.7 (0.90)	40.7
UCCSD	31.0 (1.80)	4.0 (3.00)	27.0
UCCSD(T)	24.4 (1.90)	2.8 (3.00)	21.6
CISDT	79.7 (3.00)	24.9 (0.90)	54.8
CCSDT	17.3 (2.00)	-14.2 (3.00)	31.5
CISDTQ	18.9 (2.40)	2.3 (0.90)	16.6

atom in BH, HF, and CH_4.[77] From the five energy points reported by Olsen co-workers [75] for the symmetric dissociation of H_2O (breaking two O–H bonds), the NPE is around 11

kcal/mol. Thus C_2 is a much more challenging case for CCSD than H_2O double dissociation. On the other hand, as one would expect, the non-parallelity observed here for C_2 is less severe than that seen for N_2.[76, 108, 141, 146]

As mentioned above, by the time the internuclear distances reaches 1.6 Å, the determinants $|(core)2\sigma_g^2 2\sigma_u^2 1\pi_x^2 3\sigma_g^2\rangle$ and $|(core)2\sigma_g^2 2\sigma_u^2 1\pi_y^2 3\sigma_g^2\rangle$ become very important to the ground-state FCI wave function. Choosing one or the other of these degenerate determinants leads to a symmetry-broken wave function with a lower RHF energy than that of the $|(core)2\sigma_g^2 2\sigma_u^2 1\pi_x^2 1\pi_y^2\rangle$ determinant for distances of 1.20 Å or greater. The results of using such a symmetry-broken reference function are displayed in Figure 5.4.2. CCSD and CCSD(T) results are greatly improved in the intermediate region between 1.6 and 2.6 Å. However, energies for the correlated wave functions are much worse near equilibrium, and ultimately they display the same shortcomings at large distances as in Figure 6.

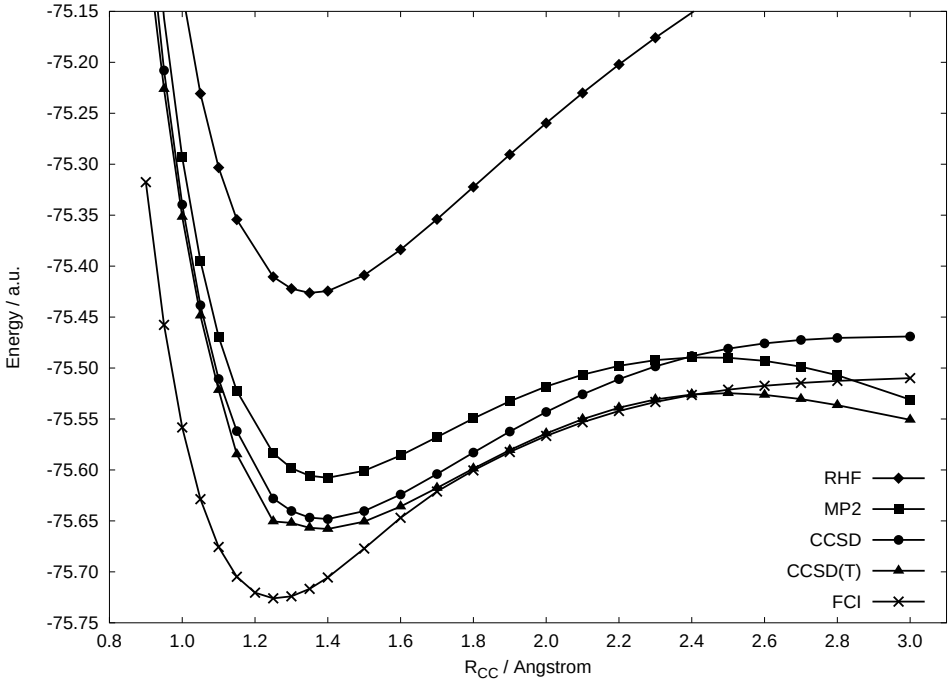

Figure 8: Potential energy curves for X $^1\Sigma_g^+$ C_2 in a 6-31G* basis set using various approximate correlation methods with a $\cdots 1\pi_x^2 1\pi_y^2 \rightarrow \cdots 3\sigma_g^2 1\pi_y^2$ symmetry-broken RHF reference.

Of course better results are to be expected for methods based on UHF references, which

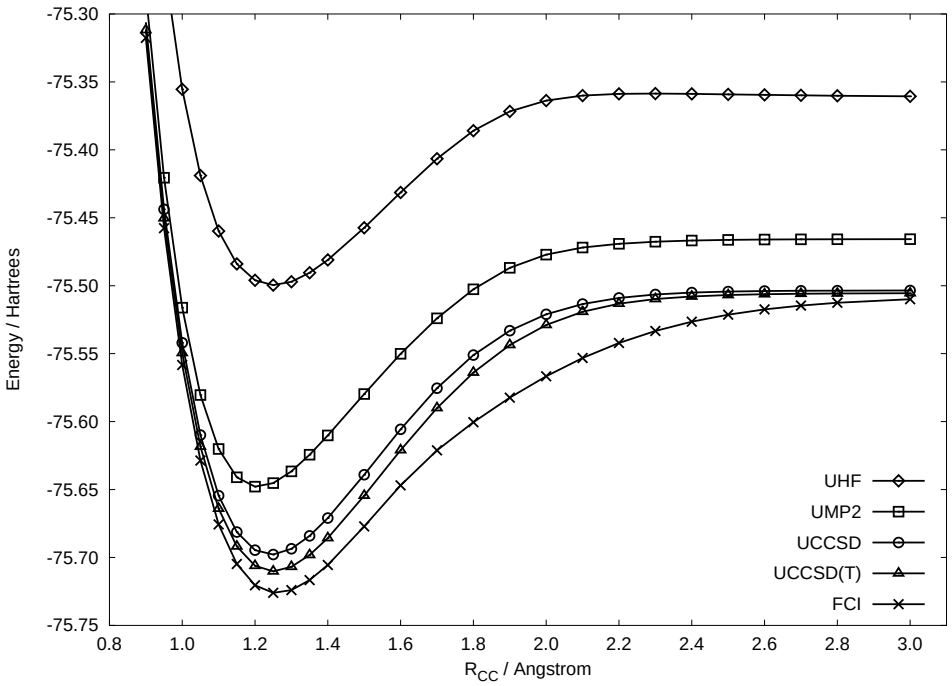

Figure 9: Potential energy curves for X $^1\Sigma_g^+$ C_2 in a 6-31G* basis set using various approximate correlation methods with a UHF reference.

at least give qualitatively correct potential energy curves for bond-breaking reactions. UHF-based results are presented in Figure 9. None of the potential energy curves diverge, but all of them rise too rapidly in the intermediate region compared to the exact FCI curve; they have difficulty in properly modeling the interaction between the two diabatic states which give rise to the adiabatic X $^1\Sigma_g^+$ and B$'$ $^1\Sigma_g^+$ states.

Error curves for UHF-based methods are plotted in Figure 10, which shows disappointingly large non-parallelities. While the restricted methods tend to have the greatest difficulty at large distances, unrestricted methods have more trouble in the intermediate region. Errors near equilibrium and the dissociation limit are actually quite small for UCCSD and UCCSD(T), but they become as large as 31 and 24 kcal/mol, respectively, at intermediate distances. The non-parallelity errors of these two methods are 27 and 22 kcal/mol, respectively. For C_2, then, the non-parallelity error of UCCSD is not improved over that of CCSD, unlike the results found for breaking bonds to hydrogen, where the NPE fell from 8-13 to 5-6 kcal/mol [77]. In H_2O, the NPE is about 9 kcal/mol for UCCSD versus 11 kcal/mol for

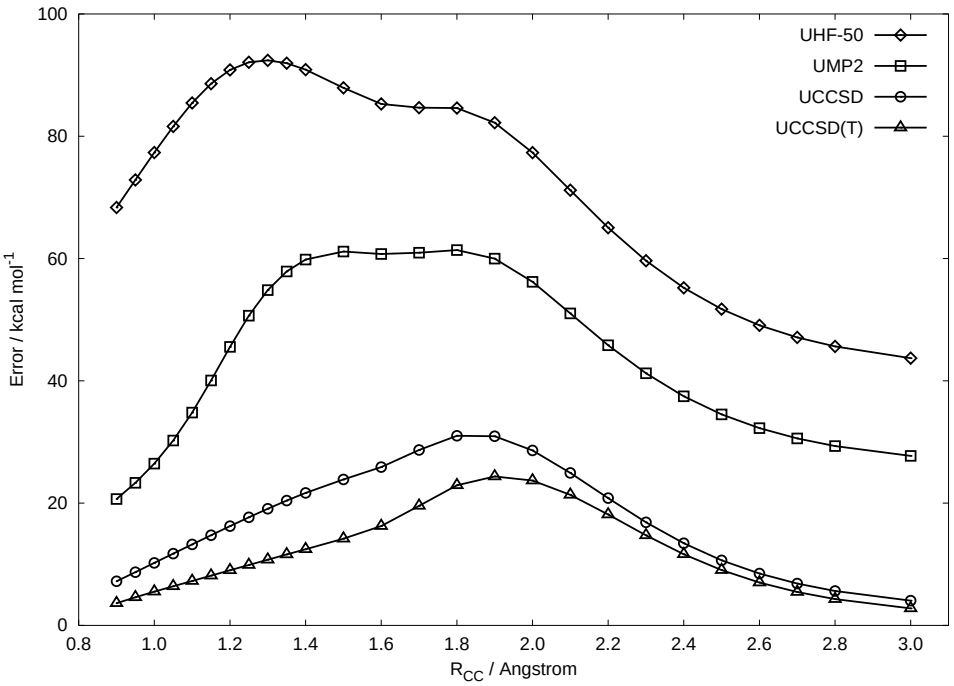

Figure 10: Errors in potential energies for X $^1\Sigma_g^+$ C_2 using various approximate correlation methods with a UHF reference and a 6-31G* basis. UHF errors have been shifted down by 50 kcal/mol.

CCSD [75].

5.4.3 Highly-correlated single-reference results for the X $^1\Sigma_g^+$ state

Given the disappointing performance of these otherwise reliable single-reference theoretical methods, one might ask what approximations are sufficient to describe these potential energy curves correctly. The multireference methods discussed in the introduction are designed to handle difficult problems like this, as are some improved single-reference methods,[139, 140] and a variety of these will be presented separately. However, in the context of the "black box" single-reference methods being discussed here, one can certainly go to higher levels of correlation, although it is not typically possible to do so for larger molecules because of the extreme increase in computational cost.

Figure 11 presents potential energy curves for the ground state using coupled-cluster theory will full, iterative triples (CCSDT) and configuration interaction through triples (CISDT) and quadruples (CISDTQ). The CISDT results are very poor, and they are even

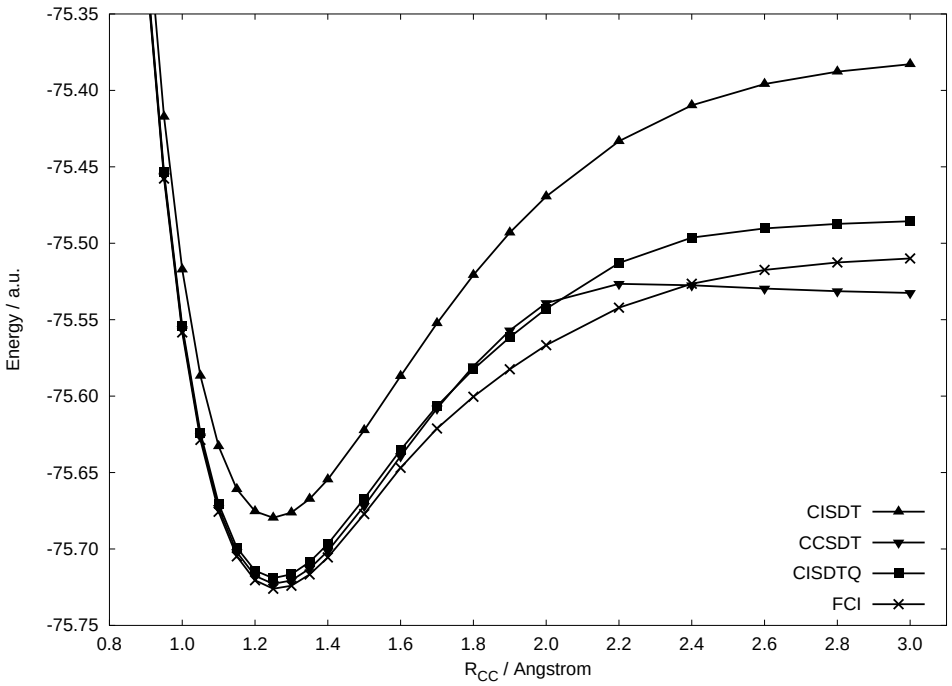

Figure 11: Potential energy curves for X $^1\Sigma_g^+$ C_2 in a 6-31G* basis set using highly correlated methods based on an RHF reference.

worse than CCSD at large distances. This is due to the disconnected quadruple substitutions (included in CCSD via the $\hat{T}_2^2$ term) being more important than triples. By including such disconnected terms, CCSDT performs much better than CISDT. At large distances, the CCSDT curve turns over and drops below the FCI energy, but the failure is not as dramatic as for CCSD(T). The variational CISDTQ method gives a curve which always stays above the FCI curve, but the error is large near dissociation (∼16-19 kcal/mol), even though it is small near equilibrium (∼ 4 kcal/mol). The non-parallelity error of CISDTQ for this double-bond breaking example, 17 kcal/mol, is similar to the 21 kcal/mol non-parallelity error of CISD for a single-bond example (BH in an aug-cc-pVQZ basis) [147]. It seems that in the single-reference truncated CI framework, two excitation levels are required for each broken bond to achieve semi-quantitative results, and an additional two excitation levels are required for quantitative results. If we use a CI wave function including all determinants through 5-fold excitations—one excitation level short of the 6-fold excitations we postulate are necessary for accurate results—we find a non-parallelity of 9 kcal/mol

between the energies at $R = 0.9$ and 2.8 Å, which is only moderately reduced from a value of 14 kcal/mol for CISDTQ. This is consistent with the modest improvement afforded by odd-numbered excitation levels in CI noted in previous studies.[75, 148]

5.4.4 Single-reference results for the B $^1\Delta_g$ and B′ $^1\Sigma_g^+$ states

Potential energy curves computed using EOM-CCSD, CISDT, and CISDTQ are presented for the B $^1\Delta_g$ and B′ $^1\Sigma_g^+$ states in Figures 12 and 13, respectively. The most commonly used excited state methods, configuration interaction singles or time-dependent density functional theory, are completely inapplicable to these states of C_2 because of the critical importance of double substitutions. EOM-CCSD provides highly reliable results for singly-excited states, and less reliable results for doubly-excited states, when the underlying CCSD is appropriate. In this case, the CCSD ground-state curves are rather poor, but they do not fail catastrophically. The figures show that the EOM-CCSD curves for the B $^1\Delta_g$ and B′ $^1\Sigma_g^+$ states lie much higher in energy than the FCI curves except at very short bond lengths. Although they roughly parallel the shape of the FCI curves at intermediate to large distances, neither could be considered a useful approximation to the FCI results. As for the ground state, CISDT falls far short of the FCI due to the importance of disconnected quadruples. CISDTQ is qualitatively correct for both curves, but quantitatively poor at large distances. The non-parallelity errors for CISDTQ are 23-24 kcal/mol for the two excited states. Modeling these two excited states is thus a great challenge for single-reference excited state theories.

5.5 Conclusions

We have presented FCI benchmark potential energy curves for the X $^1\Sigma_g^+$, B $^1\Delta_g$, and B′ $^1\Sigma_g^+$ states of C_2 using a 6-31G* basis. The standard, single-reference methods MP2, CISD, CCSD, and CCSD(T) have been assessed for their ability to reproduce the exact, FCI results for the ground state of C_2, using both restricted and unrestricted Hartree-Fock references. Many of the restricted methods are qualitatively incorrect, while all of the unrestricted methods are qualitatively correct. However, in no case are any of these approximations quantitatively satisfactory — the lowest non-parallelity error yielded by

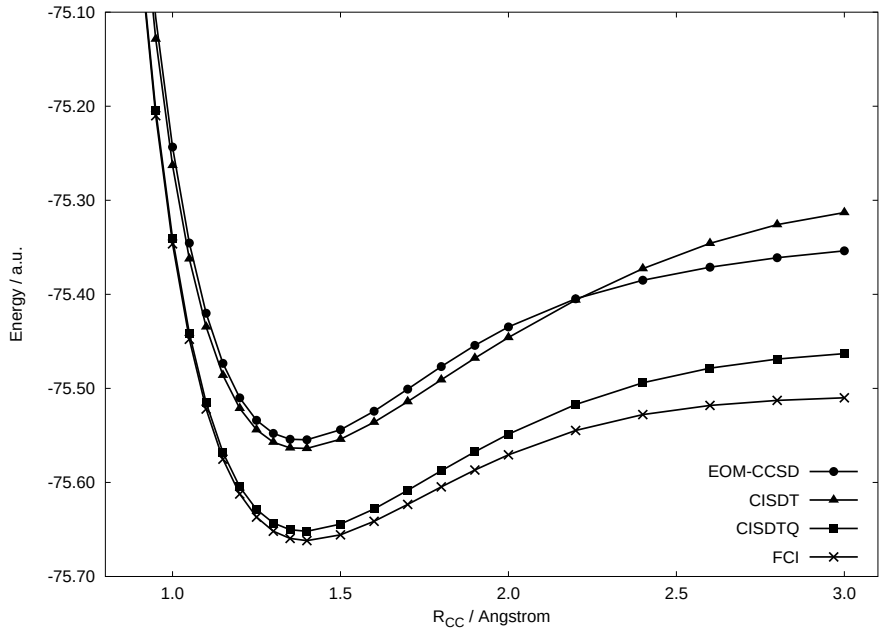

Figure 12: Potential energy curves for B $^1\Delta_g$ C_2 in a 6-31G* basis set using EOM-CCSD and highly correlated methods based on an RHF reference.

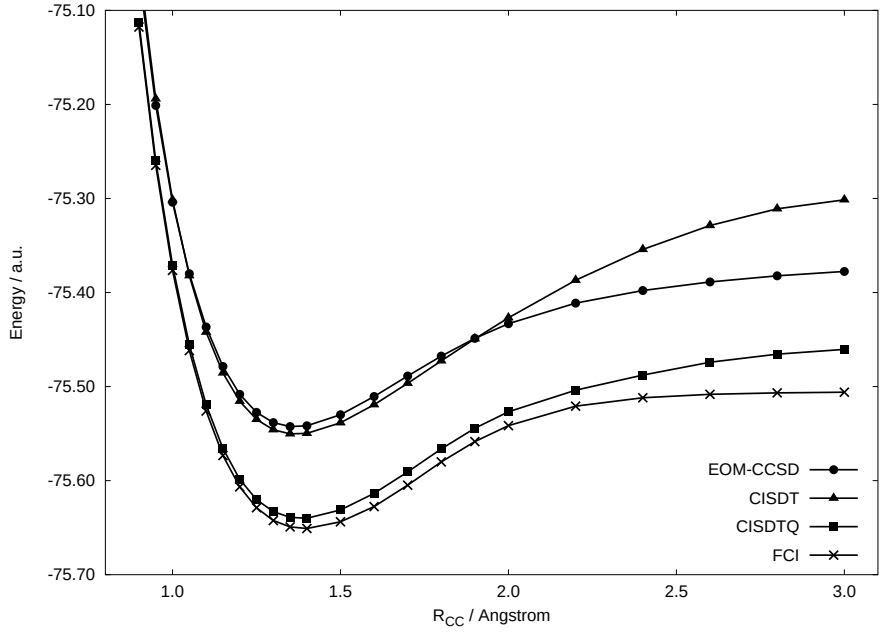

Figure 13: Potential energy curves for B$'$ $^1\Sigma_g^+$ C_2 in a 6-31G* basis set using EOM-CCSD and highly correlated methods based on an RHF reference.

any of these methods for the ground state is 22 kcal /mol, for the UCCSD(T) method. Even iterative triples in full CCSDT or triples and quadruples in CISDTQ are insufficient

to achieve quantitatively reliable results, and it appears that some inclusion of sextuple excitations is necessary. The excited states proved at least as difficult to model as the ground state.

The non-parallelity errors indicate that double bond breaking in C_2 is much more challenging than that in the symmetric dissociation of H_2O [75]. The great difficulty of single-reference methods in reproducing the FCI curves makes C_2 a very interesting test case for the calibration of theoretical methods designed for bond breaking and related problems. Such comparisons will be presented in forthcoming work.

CHAPTER 6

NATURAL ORBITALS

6.1 Abstract

Complete-active-space self-consistent-field (CASSCF) orbitals are computationally expensive and are sometimes difficult to converge. We assess complete-active-space configuration interaction (CASCI) in a basis of natural orbitals as a less expensive alternative to CASSCF. Natural orbitals are generated from various single-reference wave functions. The approach is applied to bond breaking in methyl fluoride and ethylene. With natural orbitals from correlated wave functions, CASCI parallels CASSCF potential curves, and coupled-cluster singles and doubles natural orbitals give non-parallelity errors of only 1-3 kcal/mol even for a very large active space in methyl fluoride or double bond breaking in ethylene.[1]

6.2 Introduction

The development of efficient and reliable theoretical methods for bond breaking is one of the frontier areas of modern quantum chemistry [7]. The zeroth-order Hartree-Fock wave function used in standard single-reference methods is inappropriate for describing the degeneracies that occur along a potential energy surface. Restricted Hartree-Fock (RHF) references contain unphysical ionic terms at the dissociation limit, while unrestricted Hartree-Fock (UHF) references feature incorrect behavior in the intermediate bond breaking region and massive spin contamination. A more appropriate zeroth-order wave function for breaking bonds is a multi-configurational self-consistent-field wave function (MCSCF) with a complete treatment of electron correlation in the active space; this is called complete-active-space self-consistent-field (CASSCF) [115] or full optimized reaction space (FORS) [116]. CASSCF is generally applicable to bond breaking when the active space is appropriately chosen [149]; however, the optimization of the molecular orbitals can become prohibitively

[1] M. L. Abrams and C. D. Sherrill, Chem. Phys. Lett. 395, 227-232 (2004).

expensive due to the repeated active space full configuration interaction computations. The efficiency of the CASSCF procedure is also dependent upon the quality of the starting orbitals. Modern algorithms have quadratic [150, 151] or near-quadratic [152] convergence in the local region, but can have difficulty locating the local region if the starting orbitals are a poor approximation to the final CASSCF orbitals [153].

One alternative to orbital optimization is to use natural orbitals (NOs) [81, 154], defined by Löwdin [81] as the orbitals that diagonalize the one-particle density matrix. Löwdin and Shull [155] showed that for a two-electron system, the basis of natural orbitals is the basis which requires the fewest configurations to achieve a given accuracy in the energy. Since then, several authors have demonstrated advantages of NOs over canonical Hartree-Fock orbitals [118, 156–160]. Jensen and co-workers [161] examined second-order Møller-Plesset perturbation theory (MP2) and showed that the eigenvalues of the one-particle density matrix (i.e., natural orbital occupation numbers) can be used to guide the selection of an active space. They also found that the MP2 natural orbitals provided a much better guess for a subsequent CASSCF computation than canonical Hartree-Fock orbitals. Pulay and Hamilton [162] made similar observations about UHF natural orbitals. Bofill and Pulay [158] went on to examine complete-active-space configuration interaction (CASCI) using UHF natural orbitals, which they labeled UNO-CAS, as an inexpensive alternative to CASSCF. Grev and Schaefer found that natural orbitals from a singles and doubles configuration interaction (CISD) wave function provided a good alternative to CASSCF orbitals for use in highly correlated multi-reference CI procedures [118]. More recently, Gordon and co-workers [159] showed that natural orbital occupation numbers generated from correlated wave functions can be used as a diagnostic for determining the multi-configurational character of a molecule.

A very recent paper by Bytautas, Ivanic, and Ruedenberg [163] shows that although CI wave functions may converge most rapidly in a natural orbital basis for two-electron systems [155], this result does not generalize to many-electron systems. They demonstrate that a split-localized basis derived from natural orbitals gives a more rapidly convergent CI expansion. This fascinating result indicates that other choices of orbitals may provide even

greater advantages than natural orbitals. Finally, it should be mentioned that Freed and co-workers have considered improved virtual orbitals as replacements for CASSCF orbitals in excited state computations [164], and Hirao and co-workers have explored canonical Hartree-Fock orbitals in CASCI computations [165].

The goal of the present work is to assess the accuracy of alternatives to CASSCF which replace the variationally optimized orbitals with natural orbitals. A preliminary computation is performed to obtain natural orbitals, and then a single CASCI computation is performed using those fixed orbitals. This is a much more demanding test than replacing CASSCF orbitals with natural orbitals in multi-reference CI or multi-reference perturbation theory, because those wave functions will be less dependent on the choice of orbitals (in the FCI limit, the wave function is invariant to orbital rotations). For large active spaces, the cost of the CASCI will dwarf that of the natural orbital computation, even when the natural orbitals are obtained from otherwise expensive methods such as coupled-cluster singles and doubles (CCSD). In such cases, if n iterations are required to converge the CASSCF orbitals, then the cost savings of the natural orbital CASCI approach will be approximately a factor of n. One disadvantage of this approach is that analytic energy gradients will be more difficult to formulate [166]; however, they have already been worked out for UNO-CAS [167].

In this work, CASCI in a basis of natural orbitals is compared to CASSCF for two prototypes of breaking a covalent bond: stretching the C–F bond in methyl fluoride involves breaking a σ bond between two heavy (non-hydrogen) atoms, and rotating ethylene around the HCCH torsion angle breaks a π bond. For methyl fluoride, we consider a very large, "one-to-one" active space that will provide a challenge for natural orbitals. Although most chemical reactions break only one bond at a time (even for nitrogen fixation [168]), as another challenging test case we also consider the simultaneous breaking of the σ and π C–C bonds in ethylene.

6.3 Theoretical Approach

A one-to-one active space is used for all three reactions: ($12e^-$/3a $3b_1$ $3b_2$ $3b_3$) for ethylene, and ($14e^-$/10a′ 4a″) for methyl fluoride, where the notation indicates (number of electrons/active orbitals per irrep). The active space includes the occupied valence orbitals plus a virtual, correlating orbital of the appropriate symmetry for each occupied valence orbital (the correlating virtual orbital usually has the same irreducible representation as the occupied orbital to which it corresponds; however, this may not be true for higher-symmetry point groups such as $D_{\infty h}$, where the σ bonding orbital is σ_g and the correlating orbital is σ_u^*). The result is an equal number of electrons and active orbitals. Our previous work suggests that when the one-to-one active space is larger than the valence space, it has a smaller non-parallelity error versus FCI for bond breaking [75, 169] (additional work along these lines is in progress). For ethylene, the one-to-one active space happens to be equivalent to the valence active space. Core orbitals were constrained to be doubly occupied during the CASCI.

The natural orbitals are generated from MP2, configuration interaction singles and doubles (CISD), and coupled-cluster singles and doubles (CCSD) wave functions based on an RHF reference. We also employed UHF references for the MP2 and CCSD wave functions, which we will denote as UMP2 and UCCSD. Natural orbital computations used unrelaxed density matrices and correlated all electrons (preliminary checks indicated little difference for relaxed densities). For unrestricted wave functions, the total density matrix (the sum of alpha and beta densities) was diagonalized, yielding a restricted set of natural orbitals. For comparison purposes, we also use UHF natural orbitals, as in the UNO-CAS method of Bofill and Pulay [158], and RHF orbitals. For RHF orbitals, the active orbitals are chosen as those with the lowest energy. For NOs, the active orbitals are those with the largest occupation numbers.

The reliability of the natural orbital CASCI compared to CASSCF is quantified by the non-parallelity error (NPE), defined as the difference between the maximum and minimum error along the potential energy curve. A non-parallelity error of zero indicates a potential energy curve exactly parallel to that of CASSCF. The basis sets used in this study are

6-31G** for methyl fluoride and DZP for ethylene [34, 106].

The potential energy curves for methyl fluoride were obtained constraining r(C–H)=2.069 a.u. and θ(HCF)=108.6° while stretching the C–F bond from 2.012 a.u. to 7.212 a.u. The potential energy curves for twisted ethylene used the constraints r(C–C)=1.330 Å, r(C–H)=1.076 Å, and θ(HCC)=121.7° while varying the HCCH torsion angle from 0° to 90°. For breaking both C–C bonds, r(C–C) was varied from 1.2 to 4.0 Å, keeping other parameters fixed. All calculations were performed with PSI 3.2 [144]. Due to length limitations, we are unable to report the total energies here; however, they are available from the authors. CASSCF energies for CH_3F and C_2H_4 are -139.23370 (at 2.612 bohr) and -78.18947 (at 0°), respectively.

6.4 Results and Discussion

6.4.1 $CH_3F \rightleftharpoons CH_3 + F$

The first test case involves breaking the C–F bond in methyl fluoride, for which the σ (7a′) and σ^* (8a′) orbitals should become degenerate at dissociation. It is well understood that the RHF wave function does not dissociate to the correct homolytic limit, $CH_3\cdot + F\cdot$. Because the orbitals are constrained to be doubly occupied, the wave function at dissociation contains contributions from both $CH_3^+ + F^-$ and $CH_3^- + F^+$. Figure 14, which shows the CASCI potential curves for CH_3F using different choices of orbitals, demonstrates that the errors in CASCI versus CASSCF are much larger when RHF orbitals are used. However, Figure 15 indicates that the RHF CASCI errors decrease with increasing distance, opposite to the behavior of the underlying RHF wave function. Grev and Schaefer [118] observed a similar phenomenon when CASSCF orbitals were replaced with RHF orbitals in multi-reference CI wave functions. This demonstrates that the behavior of the CASCI is not necessarily predictable from the behavior of the underlying wave function which generated the orbitals. The UHF-NO CASCI (i.e., UNO-CAS) curve begins at 3.4 bohr, the first point considered where the UHF orbitals are different than the RHF orbitals. This curve lies well below the RHF CASCI curve and is much closer to the CASSCF curve. However, as indicated in Figure 15, neither curve comes close to being parallel to the CASSCF curve.

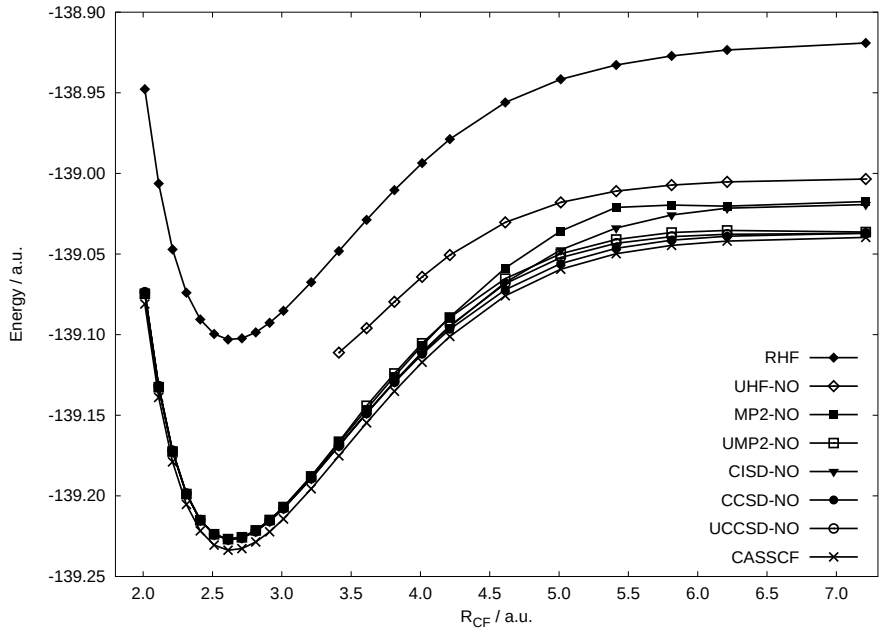

Figure 14: CASCI potential energy curves for $CH_3F \rightleftharpoons CH_3 + F$ using a 6-31G** basis set.

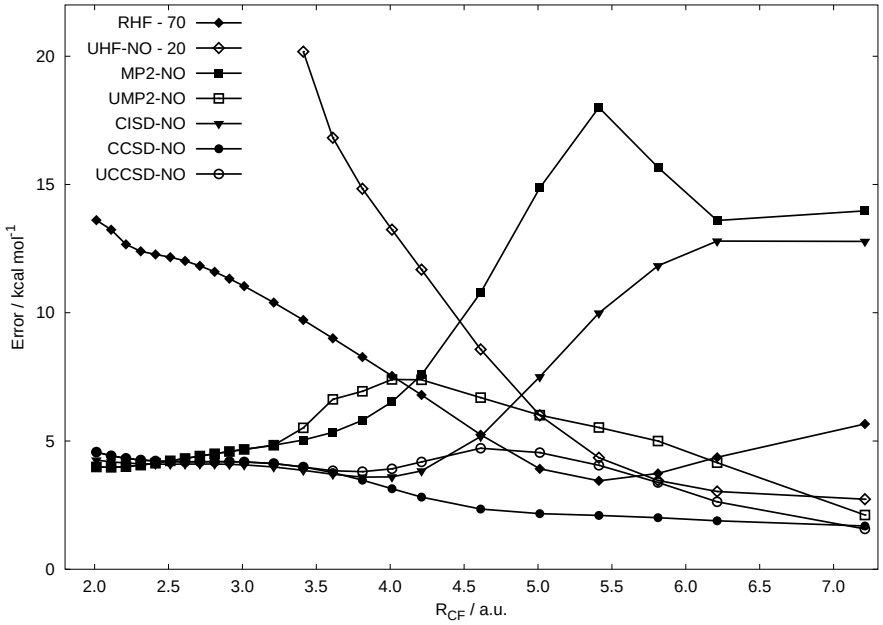

Figure 15: CASCI error curves versus CASSCF for $CH_3F \rightleftharpoons CH_3 + F$ using a 6-31G** basis set. The CASCI RHF and UHF-NO error curves are shifted by -70 and -20 kcal/mol, respectively.

Tremendous improvements are observed in Figure 14 when correlated wave functions are used to generate the natural orbitals. Indeed, all the CASCI curves based on correlated NOs lie close to the CASSCF curve, and only MP2-NO CASCI is qualitatively incorrect at large distances. The erratic behavior of the MP2-NO CASCI energy after 4.5 bohr may be attributed to the catastrophic failure of MP2 for large distances, where the energy diverges toward negative infinity. Because UHF-based MP2 dissociates single bonds qualitatively correctly, the UMP2-NO CASCI curve is much improved. Likewise, CISD, CCSD, and UCCSD curves do not exhibit qualitative failures at large distances for breaking single bonds [77], and CASCI's based on their natural orbitals perform very well.

Table 19: Maximum, minimum, and non-parallelity errors (in kcal/mol) versus CASSCF for $CH_3F \rightleftharpoons CH_3 + F$ using the 6-31G** basis set. Values in parentheses indicate the corresponding bond distance (in bohr).

Orbitals	Max error	Min error	NPE
RHF	83.61 (2.012)	73.45 (5.412)	10.16
UHF-NO	40.18 (3.412)	22.73 (7.212)	17.45
MP2-NO	18.00 (5.412)	3.97 (2.112)	14.03
UMP2-NO	7.40 (4.012)	2.11 (7.212)	5.29
CISD-NO	12.79 (6.212)	3.59 (3.812)	9.20
CCSD-NO	4.57 (2.012)	2.01 (7.212)	2.88
UCCSD-NO	4.72 (4.612)	1.57 (7.212)	3.15

Examining the errors in Figure 15, CISD-NO CASCI exhibits increasing errors at large distances, presumably due to the size-consistency error of the underlying CISD wave function. UMP2-NO CASCI has nearly flat errors at short and long distances and a peak in the error curve at intermediate distances. CCSD and UCCSD NOs give the CASCI curves most parallel to the CASSCF. A more quantitative assessment is given in Table 19, which displays the non-parallelity errors (NPEs), computed as the difference between the minimum and maximum errors versus CASSCF for the geometries considered. UMP2-NO CASCI has a smaller NPE (5.29 kcal/mol) than CISD-NO CASCI (9.20 kcal/mol), but NPEs for CCSD and UCCSD NOs are superior to both (2.88 and 3.15 kcal/mol, respectively). For CASCI computations using NOs generated from unrestricted wave functions, the maximum error

occurs at intermediate distances, where the error of the underlying unrestricted wave functions is the greatest. The minimum error, perhaps surprisingly, occurs at large distances, not small distances. It appears that the NO CASCIs have a harder time capturing the dynamical correlation near equilibrium (partially included in CASSCF due to the large active spaces tested) than the non-dynamical correlation dominant at larger distances. NPEs should be smaller than this for smaller active spaces, and in principle such large active spaces as the 14-in-14 used here are not necessary to mimic the true (FCI) potential curve if a sufficient treatment of dynamical correlation is used for orbitals beyond the active space. This general question is under investigation in our laboratory. However, we note that for a minimal active space (2-in-2), the performance of the UNO-CAS method is much better than it is in the large active space [170]. Overall, the close agreement between CASSCF and CCSD-NO CASCI across this wide range of geometries is remarkable and indicates the quality of the CCSD natural orbitals.

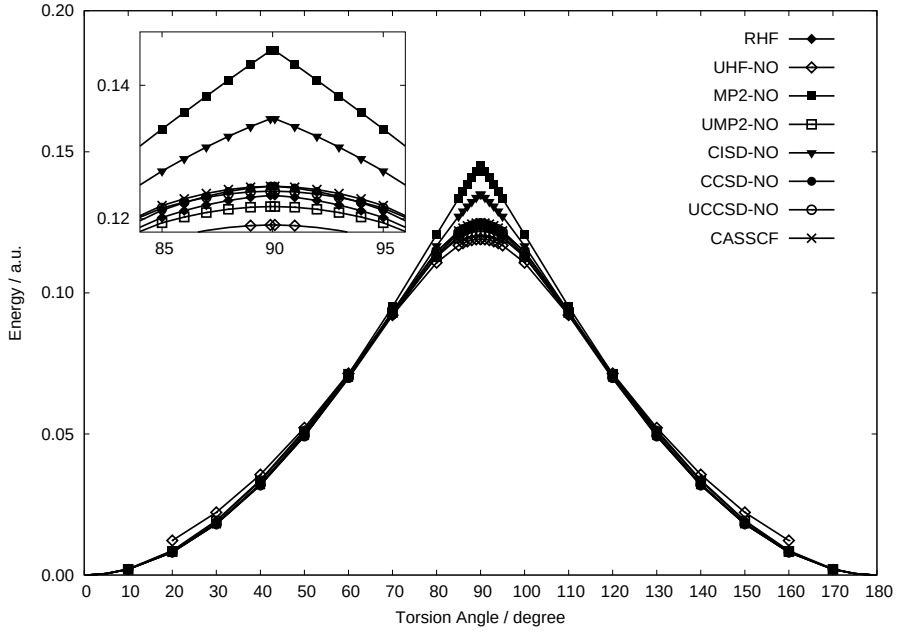

Figure 16: CASCI potential energy curves for twisted ethylene using a DZP basis set. The minimum energy at each level of theory has been set to zero. The inset contains the points nearest the cusp at 90°.

6.4.2 Twisted ethylene

The π ($1b_2$) and π^* ($2b_2$) orbitals in ethylene should become degenerate when the HCCH torsion angle is 90°. Standard spin-restricted single-reference methods are unable to handle this degeneracy properly and exhibit an unphysical cusp in the torsion potential near 90°. The UHF wave function eliminates the unphysical cusp, but it yields a torsion barrier which is far too low [106].

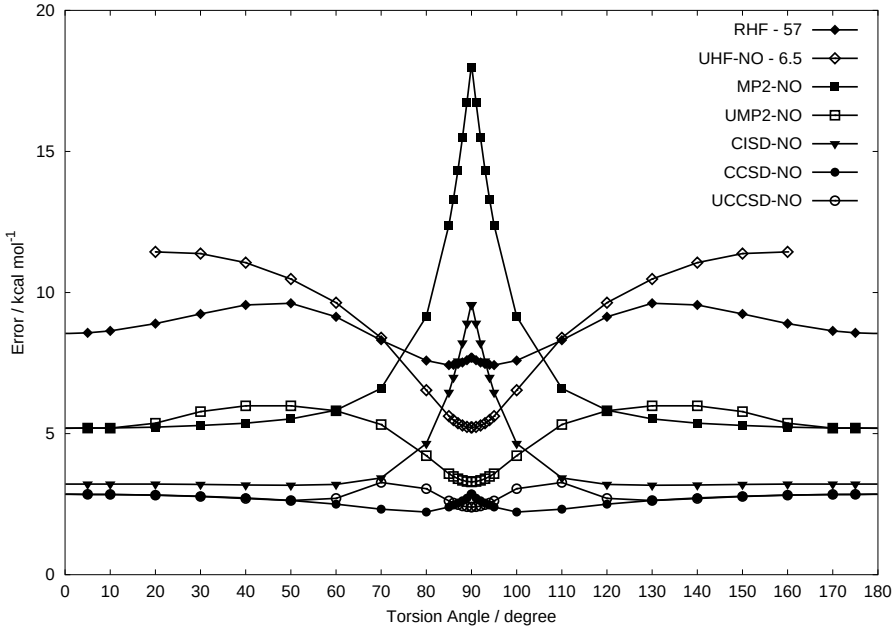

Figure 17: CASCI error curves versus CASSCF for twisted ethylene using a DZP basis set. The CASCI RHF and UHF-NO error curves are shifted by -57 and -6.5 kcal/mol, respectively.

Figure 16 shows the torsion potential with an inset of the region around the cusp, Figure 17 shows the error curves versus CASSCF, and Table 20 contains the non-parallelity errors and barrier heights. The RHF orbitals perform surprisingly well in this case. The RHF CASCI potential energy curve does not exhibit the unphysical cusp at 90°, the barrier is only 0.04 eV less than CASSCF, and the non-parallelity error is less than UMP2-NO CASCI. As seen above for methyl fluoride, however, this is not a general result for RHF CASCI.

Both the MP2-NO and CISD-NO based CASCI potential energy curves have the sharp

Table 20: Maximum, minimum, non-parallelity errors (in kcal/mol), and barrier heights (in eV) for CASCI versus CASSCF for twisted ethylene using the DZP basis set. Values in parentheses indicate the corresponding torsion angle (in °).

Orbitals	Max error	Min error	NPE	ΔE
RHF	66.62 (50.0)	64.43 (85.0)	2.19	3.36
UHF-NO	17.94 (20.0)	11.72 (90.0)	6.22	
MP2-NO	17.96 (89.9)	5.19 (0.00)	12.77	3.95
UMP2-NO	5.99 (40.0)	3.29 (90.0)	2.69	3.31
CISD-NO	9.57 (89.9)	3.17 (50.0)	6.40	3.67
CCSD-NO	2.86 (89.9)	2.22 (80.0)	0.64	3.40
UCCSD-NO	3.27 (70.0)	2.40 (90.0)	0.87	3.38
CASSCF				3.40

unphysical cusp at 90°, and the barrier heights and non-parallelity errors are much too high. These two potential energy curves mimic the potential energy curves of the underlying wave functions. UMP2-NO CASCI produces a smooth curve, and compared to MP2-NO CASCI, it reduces the non-parallelity error by more than 10 kcal/mol and the barrier height by 0.64 eV. As for the CH_3F molecule, UMP2-NO CASCI has its largest error at intermediate geometries. Both CCSD and UCCSD natural orbitals both produce CASCI potential curves which are smooth around 90°, and both have non-parallelity errors below 1 kcal/mol.

6.4.3 $C_2H_4 \rightleftharpoons 2\ CH_2$

Finally, we consider the dissociation of ethylene to form two methylenes, an example used as one of the first tests of the FORS or CASSCF approach [171]. The single-reference methods used to generate the natural orbitals are already taxed by reactions breaking a single bond, and one might suppose them incapable of providing good natural orbitals for reactions breaking a double bond. Remarkably, however, the CASCI curves in Figure 18 are not qualitatively different than those for the methyl fluoride example. Most errors in Figure 19 also appear qualitatively similar to those in Figure 15 for CH_3F. One difference is that the UMP2-NO CASCI error, rather than showing a broad peak at intermediate distances as in CH_3F, decreases gradually for C_2H_4. Conversely, the RHF CASCI error, which decreased gradually for CH_3F, shows a large increase centered around 1.6 Å. Other NOs perform quite similarly for both CH_3F and C_2H_4. Non-parallelity errors are not greatly increased for this

double bond breaking example, and for some choices of natural orbitals, they are actually smaller. The UHF NOs in particular seem to be much better in this case. As for CH_3F, the best results are seen for CASCI with natural orbitals, with NPEs of 1.91 kcal/mol for CCSD NOs and 1.77 kcal/mol for UCCSD NOs. These errors are actually smaller than those seen for methyl fluoride.

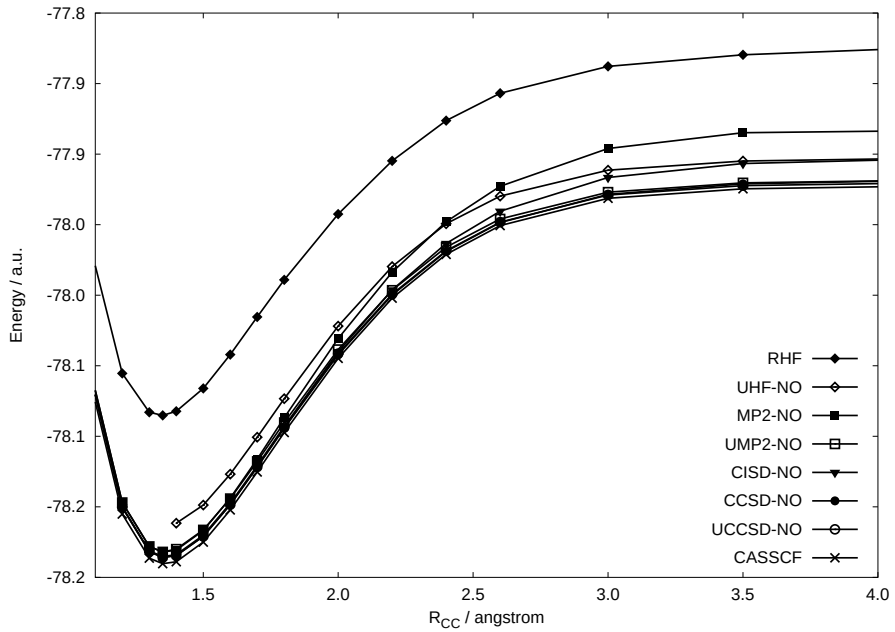

Figure 18: CASCI potential energy curves for $C_2H_4 \rightleftharpoons 2\ CH_2$ using a DZP basis set.

Table 21: Maximum, minimum, and non-parallelity errors (in kcal/mol), for CASCI versus CASSCF for $C_2H_4 \rightleftharpoons 2\ CH_2$ using the DZP basis set. Values in parentheses indicate the corresponding bond distance (in Å).

Orbitals	Max error	Min error	NPE
RHF	69.10 (1.6)	58.66 (3.0)	10.44
UHF-NO	17.14 (1.4)	12.37 (3.5)	4.77
MP2-NO	24.97 (3.5)	5.13 (1.2)	19.84
UMP2-NO	5.60 (1.4)	2.63 (4.0)	2.97
CISD-NO	11.85 (4.0)	2.74 (1.8)	9.12
CCSD-NO	3.19 (1.1)	2.40 (1.3)	1.91
UCCSD-NO	3.19 (1.1)	1.42 (4.0)	1.77

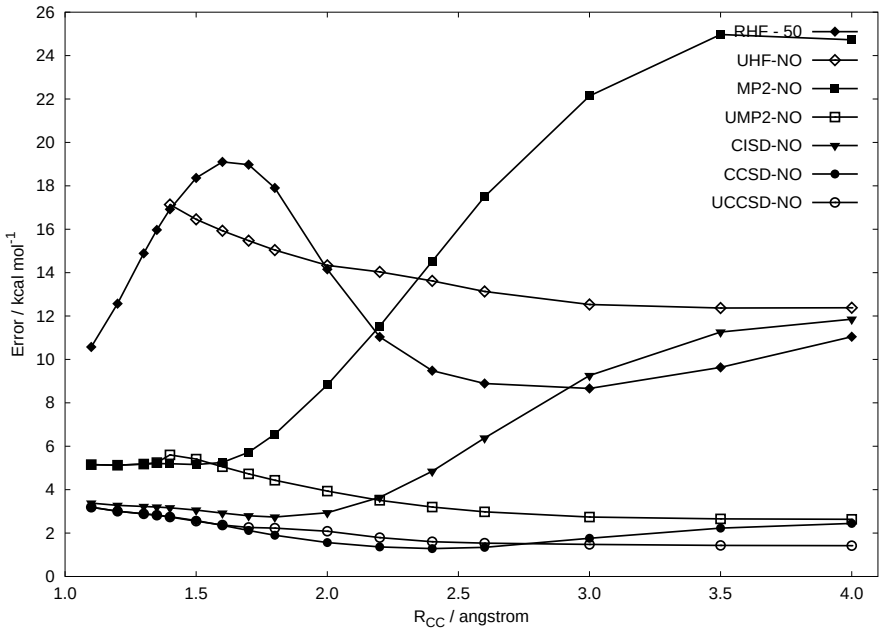

Figure 19: CASCI error curves versus CASSCF for $C_2H_4 \rightleftharpoons 2\ CH_2$ using a DZP basis set. The CASCI RHF error curve is shifted by -50 kcal/mol.

6.5 Conclusions

We have investigated natural orbitals from standard single-reference methods as alternatives to CASSCF orbitals in CASCI computations of bond-breaking reactions. When using RHF orbitals or UHF NOs, non-parallelity errors versus CASSCF can be 10 kcal/mol or larger. However, with the exception of restricted MP2, natural orbitals based on correlated wave functions generally gave CASCI energies which faithfully followed the CASSCF potential energy curves, with non-parallelity errors of a few kcal/mol. Remarkably, the reliability of this approach does not degrade significantly even when the double bond in ethylene is broken. Even if true CASSCF orbitals are desired, the present work provides the first demonstration that natural orbitals from UMP2, CCSD, or UCCSD wave functions should serve as very high quality initial guesses.

The best performance is seen for CCSD natural orbitals, for which the non-parallelity error in the CASCI is approximately 1-3 kcal/mol, even for double bond breaking and large active spaces. Although CCSD scales as N^6, this will be negligible when compared to the cost of a CASCI with a large active space, which scales factorially with the number of

active electrons and orbitals. Hence, the savings over CASSCF can be dramatic. Because multi-reference wave functions containing dynamical correlation, such as multi-reference CI, should be less sensitive to the choice of orbitals, natural orbitals should be even more successful as replacements for CASSCF orbitals in those procedures. This question will be considered in future work.

CHAPTER 7

DETERMINANT-BASED QUANTUM CHEMISTRY

7.1 Abstract

We present a determinant-based method used to formulate many-body wave functions and energy expectation values of any quantum chemical model which can be written in terms of second-quantized operators. The method is used to apply single- and multi-reference configuration interaction and coupled-cluster theories, with restricted Hartree-Fock (RHF), unrestricted Hartree-Fock (UHF), and complete-active-space self-consistent-field (CASSCF) orbitals, to the symmetric dissociation of water. Results from unrestricted state-selective multi-reference coupled-cluster theory are presented for the first time.[1]

7.2 Introduction

One of the great challenges in quantum chemistry is to develop efficient black-box methods for computing potential energy surfaces [7]. The difficulty is that different types of (quasi)-degeneracies can occur at different regions of the surface. For example, bonding and antibonding orbitals may become nearly degenerate near dissociation; a single configuration may dominate near equilibrium, but as a bond is broken a different configuration may have a weight close to that of the reference; and electronic states can become nearly degenerate at an avoided crossing or degenerate at a conical intersection. These (quasi)-degeneracies cause the electron correlation to be radically different at different regions of the surface. A quantum chemical method must be able to efficiently and, with equal accuracy, describe the electron correlation at all regions of a potential energy surface.

It is well known that restricted Hartree-Fock can dissociate to the wrong products. This is possible to correct by including electon correlation via single-reference configuration

[1] M. L. Abrams and C. D. Sherrill, Chem. Phys. Lett, accepted.

interaction, perturbation theory, or coupled-cluster theory. However, the closer the quasi-degeneracies become, i.e. the more bonds that are broken, the more excitations must be included in the single-reference method. This is computationally expensive and does not guarantee the same accuracy for all regions of a potential energy surface. If broken spin-symmetry unrestricted Hartree-Fock orbitals are used instead, the dissociative products may be correct but the wave function becomes spin-contaminated and multiple broken spin-symmetry solutions may exist. Multi-reference methods are currently the method of choice for accurately computing potential energy surfaces of small molecules [3]. However, multi-reference methods are not black-box, often relying on the ambiguous selection of reference configurations or active orbitals.

General-order algorithms were developed to ease the implementation of quantum chemical methods that include higher-order excitations, particularly in coupled-cluster theory. The first algorithms to appear in the literature used techniques from determinant-based configuration interaction [172, 173] to construct the many-body wave function and compute the expectation value of the energy [174–176]. More recently the problem was separated into two parts, the derivation of the working equations from second-quantized operators or diagrams and the efficient computer implementation of the working equations in terms of tensor contractions [177]. Kállay and co-workers were able to combine the derivation and factorization of the working equations using diagrams and the graphical techniques from determinant-based configuration interaction to efficiently perform the tensor contractions [178, 179].

Here we present our initial implementation of a determinant-based method used to implement general-order single- and multi-reference configuration interaction and coupled-cluster theories. Our efficient program for general-order configuration interaction and Møller-Plesset perturbation theory, DETCI, has been described previously [180–182]. The methods are then applied to the symmetric dissociation of water, using RHF, UHF, and CASSCF orbitals. This study represents the first report of state-selective multi-reference coupled-cluster theory using unrestricted orbitals.

7.3 Formalism

The general-order determinant-based procedure is divided into two parts, the construction of the many-body wave function and the computation of the energy expectation value. The many-body wave function, $|\Psi\rangle$, is expanded as a linear combination of Slater determinants,

$$|\Psi\rangle = \hat{\Omega}|\Phi_0\rangle = \sum_I c_I |\Phi_I\rangle \tag{7}$$

by acting on the reference determinant $|\Phi_0\rangle$ with the wave operator, $\hat{\Omega}$. The many-body wave function, in a basis of determinants, is represented by the vector $\mathbf{c}$. We use the direct configuration interaction approach [183],

$$\begin{aligned} \sigma &= \langle \Phi_I | \hat{H} | \Psi \rangle \\ &= \sum_J \langle \Phi_I | \hat{H} | \Phi_J \rangle c_J \\ &= H_{IJ} c_J \end{aligned} \tag{8}$$

to compute the energy expectation value, i.e. by acting the Hamiltonian on the many-body wave function. The resulting vector can be used to iteratively update the wave function in a variety of ways [174, 182, 184]. The difference between coupled-cluster theory [185–187] and configuration interaction theory [188] in this determinant-based method is the functional form of the wave operator.

The coupled-cluster wave operator, $\hat{\Omega}_{CC}$, is the well known exponential cluster operator,

$$\hat{\Omega}_{CC} = e^{\hat{T}} \tag{9}$$

The cluster operator is written as

$$\hat{T} = \sum_I \hat{T}_I = \sum_I t_I \hat{\tau}_I \tag{10}$$

where $\hat{\tau}_I$ is a spin-orbital excitation operator and t_I is the corresponding amplitude indexed according to the determinant $|\Phi_I\rangle$. The configuration interaction wave operator, $\hat{\Omega}_{CI}$, is defined as the first two terms, the linear terms, in the series expansion of $\hat{\Omega}_{CC}$,

$$\hat{\Omega}_{CI} = 1 + \hat{T} \tag{11}$$

The coupled-cluster wave function, $|\Psi_{CC}\rangle$, is constructed recursively with each iteration corresponding to the power of the cluster operator in the series expansion of $\hat{\Omega}_{CC}$. Three vectors, $\mathbf{c}$, $\mathbf{c}^{(n)}$, and $\mathbf{c}^{(n-1)}$, are used as follows:

$$\begin{aligned} c_I^{(n)} &= \sum_J \sum_K t_J \langle \Phi_I | \hat{\tau}_J | \Phi_K \rangle c_K^{(n-1)} \\ c_I &\leftarrow c_I^{(n)}/n! \\ c_I^{(n-1)} &\leftarrow c_I^{(n)} \end{aligned} \tag{12}$$

The first iteration, $n = 1$, includes the linear terms in $\mathbf{c}$, the second iteration, $n = 2$, includes the pair-wise disconnected terms in $\mathbf{c}$ etc., until all the connected and disconnected terms are included in $\mathbf{c}$. For a more detailed explanation see references [174] or [175].

The state-selective multi-reference wave operators use the same general form as the single-reference wave operators [186, 189, 190]. Using the label multi-reference is perhaps a misnomer, since it is an approximate version of the corresponding single-reference theory that exploits the active space concept to truncate the higher-order excitations. Either the restricted active space technique [191] or a more general active space technique [175] can be used to restrict the occupations of each subspace.

The determinant-based method is very general, making it easy to implement quantum chemical methods in terms of second-quantized operators. Unfortunately, the determinant-based general-order method is much less efficient than the standard approach. For example, the CCSD wave operator includes the disconnected doubles term, $\hat{T}_2^2$, which creates a quadruply-excited determinant when it acts on the reference determinant,

$$\hat{T}_2^2 |\Phi_0\rangle = c_Q |\Phi_Q\rangle \tag{13}$$

The general-order coupled-cluster wave function vector is expanded in a basis of determinants including up to m+2-fold excitations, while the configuration interaction wave

function vector includes only up to m-fold excitations, where m is the maximum level of excitation in the cluster operator. The Hamiltonian can connect a quadruply-excited determinant to a doubly-excited determinant, via Slater-Condon Rules,

$$\sigma_D = H_{DQ} c_Q \tag{14}$$

therefore the determinant expansion of the Hamiltonian would also include up to m+2-fold excitations. The determinant-based general-order representation of $|\Psi_{CC}\rangle$ results in a formal scaling of $O^{m+2}V^{m+2}$, where O indicates the number of occupied orbitals and V the number of virtual orbitals. The tensor formulation with factorized intermediates scales as $O^m V^{m+2}$ [177]. Two general-order coupled-cluster algorithms have been reformulated to have the correct formal scaling [178, 192].

7.4 Computational Details

The energy of H_2O in a 6-31G basis set is computed at three points along the symmetric dissociation potential energy curve. The O-H bond distances are 0.967 Å, 1.934 Å, and 2.901 Å, corresponding to R_e, $2R_e$, and $3R_e$ in Tables 22, 23, 24, and 25, with the H-O-H bond angle fixed at 107.6°. The RHF, UHF, and CASSCF(4e-/$2a_1 2b_2$) calculations were performed using PSI 3.2 [144]. All general-order single- and multi-reference configuration interaction and coupled-cluster calculations were performed with DBQC [193], using orbitals and transformed molecular integrals from PSI 3.2. The $1a_1$ core orbital was rotated during the UHF and CASSCF orbital optimizations, but was constrained to remain doubly occupied in all the correlated calculations. Therefore, calculations using UHF or CASSCF orbitals introduce an error on the order of $10^{-5} - 10^{-6}$ hartree compared to FCI computations in which the $1a_1$ core orbital is frozen in its RHF form. This error is negligible for the results presented in this study.

7.5 Results

In Tables 22, 23, 24, and 25 we report the total energy, error from full CI (FCI), and the non-parallelity errors (NPEs) for the single- and multi-reference methods. NPE is defined

as the difference between the maximum and minimum error from FCI. Though the results are for a small basis set, there are several interesting observation to be made. Since this test case has been investigated previously using larger basis sets [179, 194], we can gauge the basis set dependence of our observations.

Table 22: The FCI total energies, the error from FCI for the three reference methods, and the non-parallelity error in hartrees (kcal/mol in parentheses).

Method	R_e	$2R_e$	$3R_e$	NPE	
RHF	0.136672	0.295881	0.505063	0.368392	(231.17)
UHF	0.136672	0.095935	0.099397	0.040736	(25.56)
CAS(4,4)	0.083260	0.068682	0.060733	0.022528	(14.14)
FCI	-76.121174	-75.876474	-75.836657		

Table 23: The error from FCI and the non-parallelity error in hartrees (kcal/mol in parentheses).

Method	R_e	$2R_e$	$3R_e$	NPE	
CISD	0.006858	0.055476	0.178951	0.172093	(107.99)
CISDT	0.005854	0.045535	0.158994	0.153140	(96.10)
CISDTQ	0.000175	0.003742	0.010179	0.010004	(6.28)
CISDTQP	0.000103	0.001522	0.004136	0.004032	(2.53)
CISDTQPH	0.000001	0.000039	0.000101	0.000100	(0.06)
UCISD	0.006858	0.022895	0.002173	0.020721	(13.00)
UCISDT	0.005854	0.019551	0.001821	0.017730	(11.13)
UCISDTQ	0.000175	0.006567	0.000588	0.006393	(4.01)
UCISDTQP	0.000103	0.003375	0.000437	0.003272	(2.05)
UCISDTQPH	0.000001	0.000147	0.000160	0.000159	(0.10)

The single-reference methods using RHF orbitals exhibit similar patterns to previously published results. To achieve an NPE less than 0.001 hartree (0.6 kcal/mol), configuration interaction and coupled-cluster theory require up to hextuple and pentuple excitations, respectively. Most of the error in the coupled-cluster results occur because the energy drops below FCI at extended geometries. As pointed out by Paldus and co-workers [195], the disconnected terms can become unphysically large for (quasi)-degenerate configurations. To correct the disconnected terms, the appropriate higher-order connected terms should be included. It is surprising to see that at $3R_e$ the CCSDTQ energy using RHF orbitals

Table 24: The error from FCI and the non-parallelity error in hartrees (kcal/mol in parentheses).

Method	R_e	$2R_e$	$3R_e$	NPE	
CCSD	0.001545	0.009846	-0.020882	0.030728	(19.28)
CCSDT	0.000449	-0.001965	-0.040860	0.041309	(25.92)
CCSDTQ	0.000012	0.000102	-0.002782	0.002884	(1.81)
CCSDTQP	0.000003	0.000023	-0.000046	0.000069	(0.04)
UCCSD	0.001545	0.012696	0.001359	0.011337	(7.11)
UCCSDT	0.000449	0.005351	0.000390	0.004961	(3.11)
UCCSDTQ	0.000012	0.001211	0.000266	0.001199	(0.75)
UCCSDTQP	0.000003	0.000156	0.000132	0.000153	(0.10)

Table 25: The error from FCI and the non-parallelity error in hartrees (kcal/mol in parentheses).

Method	Orbitals	R_e	$2R_e$	$3R_e$	NPE	
MRCISD	RHF	0.004876	0.003943	0.004211	0.000933	(0.59)
MRCCSD	RHF	0.000867	0.000824	0.000898	0.000074	(0.05)
MRCISD	UHF	0.004876	0.005645	0.002008	0.003637	(2.28)
MRCCSD	UHF	0.000867	0.001415	0.000986	0.000548	(0.34)
MRCISD	CAS(4,4)	0.001476	0.001485	0.001250	0.000235	(0.15)
MRCCSD	CAS(4,4)	0.000291	0.000299	0.000381	0.000091	(0.06)

is approximately 2.0 kcal/mol lower than the FCI energy. When pentuple-excitations are included, the error from FCI decreases by a factor of 50, but remains below the FCI energy.

The trends do change when UHF orbitals are used. The largest error for the single-reference methods is now at $2R_e$ instead of $3R_e$ and the lower-order methods have a significantly smaller NPE. The NPE does not drop below 0.001 hartree until hextuple and pentuple excitations are included in the wave function, just as with the RHF orbitals. The UCCSDTQ NPE is only 0.8 kcal/mol; however, the method scales formally as N^{10} or O^4V^6, which precludes its application to systems with more than a few heavy atoms or a large basis set.

As expected, the multi-reference methods exhibit significantly smaller NPEs. Except for MRCISD with UHF orbitals, all the NPEs are below 0.001 hartree. The MRCCSD energy, regardless of the choice of orbitals, does not drop below the FCI energy. According

to Paldus and co-workers [195] this occurs because the MRCCSD wave operator includes the important connected pentuple and hextuple excitations required to correct the unphysically large disconnected terms from the single, double, triple, and quadruple excitations. The difference between the MRCISD and the MRCCSD NPEs is only 0.000144 hartree, significantly smaller than one might expect. The difference between these two methods is similar for the potential energy curves of HF [175] and N_2 [196, 197].

Previous studies [179, 194] using the cc-pVDZ basis set showed that MRCISD using different active spaces, including the (4,4) used in the present study, have NPEs greater than 0.001 hartree, indicating that correlation effects are somewhat more challenging for the larger basis set. An NPE of less than 1 kcal/mol was achieved, but only with the much larger one-to-one active space. It is interesting to note that CASSCF with a one-to-one active space was the only method to have an NPE of approximately 0.001 hartree for the cc-pVDZ basis set, which can only be attributed to a fortuitous cancellation of errors.

We have described our initial implementation of a general-order determinant-based method using restricted occupations of orbital subspaces. Our single- and multi-reference results for the symmetric dissociation of water show that excitations beyond quadruples are required to achieve an NPE of 0.001 millihartree (0.6 kcal/mol). Using multi-reference methods with UHF orbitals only increased the NPE compared to RHF or CASSCF orbitals. Work on developing compact CI and CC wave functions without using active spaces or orbital optimization are presently underway. We also plan to explore alternative general-order algorithms with the correct formal scaling.

CHAPTER 8

IMPORTANT CONFIGURATIONS

8.1 Abstract

Using a new general-order determinant-based program, we construct compact configuration interaction and coupled-cluster wave functions by selecting the most important configurations, by weight, from a full configuration interaction or full coupled-cluster wave function. Our results show that for the symmetric dissociation of water, chemical accuracy can be achieved across the surface with $\sim 2\%$ of the full coupled-cluster expansion compared to $\sim 10\%$ of the full configuration interaction expansion.

8.2 Introduction

Single-reference methods and algorithms in quantum chemistry have made tremendous progress in the last 20 years. As an example, a recent study [1] reported 1 kJ/mol accuracy (compared to experiment) for the computed enthalpy of formation (at 0 K) for 31 atoms and molecules. For (quasi)-degenerate molecular systems this level of accuracy, commonly referred to as "chemical accuracy," has yet to be achieved and is currently one of the primary areas of development [3, 7].

The calculation of accurate potential energy surfaces is a particularly challenging problem. As a chemical bond is stretched to dissociation, triple, quadruple, and even higher excitations can become important. Including an entire set of excitations beyond doubles is too expensive for all but the smallest molecules and basis sets and, for non-variational methods, the energy can drop significantly below the full CI energy.

Two principal techniques are currently used to include subsets of these higher-order excitations. The first is to divide the orbitals into subspaces and restrict the occupations of each subspace. This has been applied in configuration interaction [175, 182, 188, 191, 198, 199], perturbation theory [199–202], and coupled-cluster theory [189, 190, 203, 204]. The

other, primarily used in configuration interaction, is to select a configuration by estimating its contribution to the energy or to the wave function [182, 188, 205–208].

The recent development of general-order coupled-cluster programs has extended the ability to develop and test coupled-cluster methods with higher-order excitations [175, 179, 209]. As new techniques are developed to construct compact coupled-cluster wave functions for accurately describing (quasi)-degenerate molecular systems, it is important to determine how compact such a coupled-cluster wave function can be. Additionally, we desire a wave function which is size extensive and avoids a non-variational divergence of the energy.

In the present study, we construct truncated CI and, for the first time, truncated CC wave functions from an a posteriori selection, by weight, of the configurations from a full CI (FCI) and full CC (FCC) wave function, respectively. By selecting the most important configurations from a full expansion, we can gauge the limit of the accuracy and the size of the expansion to expect from future work on developing a priori selection schemes. A similar procedure was applied by Ivanic and Reudenberg [210] for constructing CI wave functions. We apply this technique for CI and CC wave functions to three points corresponding to the symmetric dissociation of water, where comparisons can be made to previous studies of single and multi-reference methods [179, 194, 209].

8.3 Computational Details

Three points on the potential energy curve corresponding to the symmetric dissociation of H_2O are obtained by fixing the H-O-H bond angle to 107.6° and simultaneously stretching the O-H bonds from R_e to $2R_e$ and $3R_e$, with R_e equal to 0.967 Å. The 6-31G basis set was used and the $1a_1$ core orbital was frozen. All configuration interaction and coupled-cluster calculations were performed with DBQC [193], using the orbitals and transformed molecular integrals from PSI 3.2 [125].

The procedure used to construct the truncated CI and CC wave functions begins with solving for the FCI and FCC wave functions. At this stage, we do not use an a priori selection scheme to efficiently include configurations. Instead we use the best possible wave function to determine, in principle, how accurate a truncated wave function can be. The

importance of a CI or CC coefficient is determined by its weight in the FCI or FCC wave function vector. Of course, FCI and FCC are equivalent procedures, but they use different parameterizations of the wave function. The coefficients with a weight greater than a chosen threshold ($10^{-3.0}$, $10^{-3.5}$, $10^{-4.0}$) become the basis for the expansion of the truncated wave function. This selection procedure does not guarantee that the truncated expansion will yield an eigenfunction of $\hat{S}^2$. Additional coefficients, which have a weight lower than the threshold, are added to maintain the proper spin-symmetry of the truncated wave function. Finally, the truncated wave function is re-optimized.

8.4 Results

Tables 26 and 27 show the energy of the truncated CI and CC wave functions, respectively, for the three thresholds and the three points corresponding to the symmetric dissociation of H_2O. We also report the non-parallelity error (NPE), defined as the difference between the maximum error and the minimum error from FCI along the potential energy curve. Using a threshold of $10^{-3.0}$, the CI NPE is nearly 2.5 times as large as the CC NPE. The CI and CC errors from FCI are nearly the same at $2R_e$ and $3R_e$, but at R_e the CC error from FCI is much smaller. As the threshold is decreased and more configurations are included in the truncated wave function, the difference between CI and CC becomes smaller. Using a threshold of $10^{-4.0}$, the CI NPE is actually smaller than the CC NPE.

Table 26: Truncated CI total energy, error from FCI, and non-parallelity error (in hartree, kcal/mol in parentheses) using one of three thresholds for three O-H bond distances (in Å) corresponding to the symmetric dissociation of H_2O. The FCI energy is -76.121174 at R_e, -75.876474 at $2R_e$, and -75.836657 at $3R_e$.

Threshold	R_{OH}	Energy	Error	NPE
$10^{-3.0}$	0.967	-76.117703	0.003471	0.002522 (1.58)
$10^{-3.0}$	1.934	-75.874554	0.001920	
$10^{-3.0}$	2.901	-75.835708	0.000949	
$10^{-3.5}$	0.967	-76.120624	0.000550	0.000398 (0.25)
$10^{-3.5}$	1.934	-75.876069	0.000405	
$10^{-3.5}$	2.901	-75.836505	0.000152	
$10^{-4.0}$	0.967	-76.121049	0.000126	0.000102 (0.06)
$10^{-4.0}$	1.934	-75.876416	0.000058	
$10^{-4.0}$	2.901	-75.836634	0.000023	

Table 27: Truncated CC total energy, error from FCI, and non-parallelity error (in hartree, kcal/mol in parentheses) using one of three thresholds for three O-H bond distances (in Å) corresponding to the symmetric dissociation of H_2O. The FCI energy is -76.121174 at R_e, -75.876474 at $2R_e$, and -75.836657 at $3R_e$.

Threshold	R_{OH}	Energy	Error	NPE
$10^{-3.0}$	0.967	-76.119863	0.001311	0.000952 (0.60)
$10^{-3.0}$	1.934	-75.874711	0.001763	
$10^{-3.0}$	2.901	-75.835846	0.000811	
$10^{-3.5}$	0.967	-76.120634	0.000540	0.000377 (0.24)
$10^{-3.5}$	1.934	-75.876041	0.000433	
$10^{-3.5}$	2.901	-75.836494	0.000163	
$10^{-4.0}$	0.967	-76.120983	0.000191	0.000157 (0.10)
$10^{-4.0}$	1.934	-75.876379	0.000095	
$10^{-4.0}$	2.901	-75.836622	0.000035	

Tables 28 and 29 show the number of configurations of each excitation level included in the truncated CI and CC wave functions, respectively, for the three thresholds. In general, as the two O-H bonds are stretched the number of configurations increases and the maximum excitation level increases. The truncated CI wave function, using a threshold of $10^{-3.0}$, includes up to quadruple excitations at R_e, hextuple excitations at $2R_e$, and octuple excitations at $3R_e$. A similar distribution of configurations for the symmetric dissociation of water was report by Greer [208] when the CI configurations were randomly selected. The truncated CC wave function, using the same threshold, only includes up to triple excitations at R_e, quadruple excitations at $2R_e$, and pentuple excitations at $3R_e$. The truncated CC wave function includes far fewer configurations for all three thresholds. An NPE of less than 1 millihartree is achieved with approximately 2% of the CC configurations, while more than 10% of the CI configurations would be required to achieve the same accuracy.

Table 28: Number of configurations according to excitation level for the truncated CI wave function using one of three thresholds for three O-H bond distances (in Å) corresponding to the symmetric dissociation of H_2O. Also given are the total number of configurations and the percentage of the full configuration space. Total number of configurations in the full wave function according to excitation level: 24, 384, 2792, 10550, 19936, 18680, 7840, 1234; for a total of 61440.

Threshold	R_{OH}	$\|\Phi_1\rangle$	$\|\Phi_2\rangle$	$\|\Phi_3\rangle$	$\|\Phi_4\rangle$	$\|\Phi_5\rangle$	$\|\Phi_6\rangle$	$\|\Phi_7\rangle$	$\|\Phi_8\rangle$	Total	%Full
$10^{-3.0}$	0.967	20	370	38	700	0	0	0	0	1128	1.8
$10^{-3.0}$	1.934	22	342	822	1454	322	102	0	0	3064	5.0
$10^{-3.0}$	2.901	24	342	1152	1910	1034	259	24	1	4746	7.7
$10^{-3.5}$	0.967	22	374	1352	3805	4	0	0	0	5557	9.0
$10^{-3.5}$	1.934	22	376	1596	3454	1188	433	42	1	7112	11.6
$10^{-3.5}$	2.901	24	362	1866	3610	3204	1190	176	11	10443	17.0
$10^{-4.0}$	0.967	24	378	2302	6545	358	750	0	0	10357	16.9
$10^{-4.0}$	1.934	22	379	2338	6190	4026	1845	228	11	15039	24.5
$10^{-4.0}$	2.901	24	374	2270	5743	6102	2995	620	39	18167	29.6

Table 29: Number of configurations according to excitation level for the truncated CC wave function using one of three thresholds for three O-H bond distances (in Å) corresponding to the symmetric dissociation of H_2O. Also given are the total number of configurations and the percentage of the full configuration space. Total number of configurations in the full wave function according to excitation level: 24, 384, 2792, 10550, 19936, 18680, 7840, 1234; for a total of 61440.

Threshold	R_{OH}	$\|\Phi_1\rangle$	$\|\Phi_2\rangle$	$\|\Phi_3\rangle$	$\|\Phi_4\rangle$	$\|\Phi_5\rangle$	$\|\Phi_6\rangle$	$\|\Phi_7\rangle$	$\|\Phi_8\rangle$	Total	%Full
$10^{-3.0}$	0.967	20	372	70	0	0	0	0	0	462	0.8
$10^{-3.0}$	1.934	22	332	588	64	0	0	0	0	1006	1.6
$10^{-3.0}$	2.901	24	332	718	161	116	0	0	0	1351	2.2
$10^{-3.5}$	0.967	22	374	1408	3	0	0	0	0	1807	2.9
$10^{-3.5}$	1.934	22	376	1354	582	56	0	0	0	2390	3.9
$10^{-3.5}$	2.901	24	360	1244	984	350	55	0	0	3017	4.9
$10^{-4.0}$	0.967	24	378	2302	585	0	0	0	0	3289	5.4
$10^{-4.0}$	1.934	22	379	2180	2331	406	7	0	0	5325	8.7
$10^{-4.0}$	2.901	24	374	2058	3432	1196	613	3	0	7699	12.5

When (quasi)-degeneracies are present, the disconnected terms in the exponential wave function can become unphysically large thus causing the CC energy to drop below the FCI energy [195]. This can occur even if the orbitals are optimized in conjunction with the coupled-cluster amplitudes [211]. The CC energy can be corrected if higher-order connected terms are included in the expansion of the wave function. Other researchers have explored amplitude-corrected and energy-corrected coupled-cluster theory as a way to include the effects of higher-order connected configurations [139, 212]. These methods produce accurate potential energy curves, but at the expense of losing rigorous size-extensivity. The truncated CC wave functions are size-extensive and include the most important higher-order configurations. However, the a priori selection of the higher-order connected configurations could be a delicate process. For example, if the 116 pentuply-excited configurations are removed from the truncated CC wave function (at $3R_e$ and a threshold of $10^{-3.0}$), the energy drops below the FCI energy by approximately 2.0 kcal/mol!

Selecting the important configurations by weight can, in principle, produce a truncated CC wave function that accurately describes the electron correlation at equilibrium and at stretched geometries. It is interesting to note that at equilibrium the singular value decomposition (SVD) of the triples amplitudes [213] and the truncated CC wave function from the present study produce similar absolute errors from FCI ($\sim$ 0.2-2.0 millihartree) and a similar ratio of triples amplitudes to doubles amplitudes ($\sim$ 2-20%). Since the SVD of the CC amplitudes has only been applied to doubles [214] and triples [213] amplitudes, the compressed wave functions will be unable to achieve chemical accuracy for breaking double or triple bonds.

8.5 Conclusions

We show that a compact CC wave function, including $\sim$ 2% of the full configuration space, can accurately describe the symmetric dissociation of water without the non-varitional divergence of the energy and without any type of post-Hartree-Fock orbital optimization. It is well known that the CC formalism is less dependent on the choice of orbitals than the CI formalism. Future work will investigate the dependence of the number and type of

important configurations on the choice of orbitals, and how the important configurations may be chosen a priori without knowledge of the full wave function.

CHAPTER 9

CONCLUSIONS

The ultimate goal of any computational science is to develop models that rival or even exceed the accuracy of experimental measurements. A more realistic goal would be the ability to specify a system, a property of the system, and the accuracy to which the property should be computed, and only be limited by computational resources at hand. The introduction of coupled-cluster theory to quantum chemistry was a major step toward predictive accuracy for molecules with more than a few atoms [215]. The hierarchy of coupled-cluster methods and systematically constructed basis sets have reached unprecedented accuracy often equal to or exceeding experimental measurements [1, 2]. However, if (quasi)-degeneracies are present in the molecular wave function, the hierarchy of coupled-cluster methods greatly deteriorates. Systematically increasing the excitation level is not guaranteed to provide a more accurate result, and often does not, until the excitation level is so high that the computation becomes intractable. The general-order, general-configuration selection procedure, presented in this thesis, can extend the application of the coupled-cluster method to predictive accuracy for (quasi)-degenerate molecular wave functions.

If the general-order, general-configuration selection coupled-cluster method is to become a practical method in quantum chemistry, two principal theoretical and computational advances must be made. First, an a priori selection scheme, similar to one developed for truncating the full configuration interaction wave function [182, 188, 205–208], must be developed and calibrated. The procedure should include an appropriate starting list of configurations, a method to estimate the contribution an interacting configuration will have on the energy and the wave function, and the interacting configurations must be chosen in an efficient and systematic way. Recent insights from amplitude-corrected and energy-corrected coupled-cluster methods [139, 212] should help guide this development. Second, the determinant-based algorithm presented in Chapter 7 is too computationally expensive.

More efficient [correct $O(N^{2m+2})$ scaling with N orbitals and highest excitation level m] general-order coupled-cluster algorithms have been developed for the standard hierarchy (CCSD, CCSDT, CCSDTQ, ...) [177, 178] and for restricted active spaces [179], but would need to be significantly altered to handle general lists of configurations.

With current computational facilities, the scaling of traditional coupled cluster methods [CCSD $O(N^6)$ and CCSD(T) $O(N^7)$] will rapidly become the bottleneck for systems containing more than a dozen atoms. An intelligent selection scheme will naturally exclude configurations corresponding to spatially-distant weakly-interacting electrons because the contribution to the energy and the wave function will be small. However, the enormous number of configurations must first be tested in order to be excluded, resulting in unreasonable computation times. An alternative approach for large molecules would be to transform the molecular orbitals to a local basis [216–219], which would make it possible to exclude a large percentage of the configurations at the beginning of the calculation, making the method much more efficient.

The general-order, general-configuration selection coupled-cluster method can be extended to anything that can be done with traditional coupled-cluster theory [185–187], including ground-state and excited-state energies and properties. The standard approach to excitation energies and excited-state properties is the equation-of-motion or linear-response method [220, 221]. It involves the diagonalization of the non-Hermitian similarity transformed Hamiltonian ($e^{-\hat{T}}\hat{H}e^{\hat{T}}$) in a space spanned by a linear set of determinants. The linear set of determinants, as in configuration interaction, lead to size-extensivity errors. Equation-of-motion coupled-cluster theory has been so successful because the excitation energy is size-intensive, meaning the excitation energy of the non-interacting subsystems is obtained as the excitation energy of the supermolecule [222]. Furthermore, the coupled-cluster bra state ($\langle\Psi_{CC}|$), which is required for molecular properties, contains disconnected terms resulting in size-extensivity errors. The standard coupled-cluster method can be modified to generate a size-extensive bra state [223, 224]; however, the increased computational scaling has prohibited its application to larger molecular systems. The application of coupled-cluster theory to general open-shell molecules has met similar difficulties [186].

The most formidable hurdle has been the number of equations resulting from the extension of the Cambell-Baker-Hausdorff expansion of the similarity transformed Hamiltonian from four commutators to eight. Coupled-cluster theory still requires new formal and computational advances to enable the routine computation of any molecular system and any property.

Perhaps it is surprising that higher-order connected clusters must be included in the expansion of the wave function at all. The results from Chapters 7 and 8 showed that the number of higher-order clusters can be greatly reduced by an appropriately chosen active space or numerical selection criterion. But as the number and strength of (quasi)-degeneracies increases (potential energy curve of Cr_2), clusters beyond hextuples, which already scale $O(N^{14})$, will become important. There are alternative approaches to solving the electronic Schrödinger equation that are less sensitive to (quasi)-degeneracies, though less "mature" than wave function-based methods in quantum chemistry, e.g. quantum Monte Carlo [225–228], density matrix renormalization group [197, 229–231], and reduced density matrix methods [232–235]. It is possible that an alternative method could eventually provide a more accurate and efficient description of systems with strongly-interacting electronic configurations.

REFERENCES

[1] A. Tajti, A. G. C. P. G. Szalay, M. Kállay, J. Gauss, E. F. Valeev, B. A. Flowers, J. Vázquez, and J. F. Stanton, J. Chem. Phys. **121**, 11599 (2004).

[2] T. Helgaker, T. A. Ruden, P. Jorgensen, J. Olsen, and W. Klopper, J. Phys. Org. Chem. **17**, 913 (2004).

[3] K. Hirao, editor, *Recent Advances in Multireference Methods*, volume 4 of *Recent Advances in Computational Chemistry*. World Scientific, Singapore, 1999.

[4] H.-J. Werner and P. J. Knowles, J. Chem. Phys. **88**, 5803 (1988).

[5] H.-J. Werner, Mol. Phys. **89**, 645 (1996).

[6] P. Celani and H.-J. Werner, J. Chem. Phys. **112**, 5546 (2000).

[7] M. R. Hoffmann and K. G. Dyall, editors, *Low-Lying Potential Energy Surfaces*, ACS Symposium Series 828. American Chemical Society, Washington, DC, 2002.

[8] H. Köppel, W. Domcke, L. S. Cederbaum, and W. von Diessen, J. Chem. Phys. **69**, 4252 (1978).

[9] A. J. Lorquet and J. C. Lorquet, The Journal of Chemical Physics **49**, 4955 (1968).

[10] M. J. S. Dewar and H. S. Rzepa, J. Am. Chem. Soc. **99**, 7432 (1977).

[11] J. A. Pople, W. A. Lathan, and W. J. Hehre, J. Am. Chem. Soc. **93**, 808 (1971).

[12] R. J. Buenker, S. D. Peyerimhoff, and H. L. Hsu, Chem. Phys. Lett. **11**, 65 (1971).

[13] W. R. Rodwell, M. F. Guest, D. T. Clark, and D. Shuttleworth, Chem. Phys. Lett. **45**, 50 (1977).

[14] N. C. Handy, R. H. Nobes, and H.-J. Werner, Chem. Phys. Lett. **110**, 459 (1984).

[15] N. Salhi-Benachenhou, B. Engels, M.-B. Huang, and S. Lunell, Chem. Phys. **236**, 53 (1998).

[16] L. A. Eriksson, S. Lunell, and R. J. Boyd, J. Am. Chem. Soc. **115**, 6896 (1993).

[17] Y.-J. Liu and M.-B. Huang, J. Mol. Struct. (Theochem) **536**, 133 (2001).

[18] K. Toriyama and M. Okazaki, Appl. Magn. Reson. **11**, 47 (1996).

[19] K. Toriyama and M. Okazaki, Acta Chem. Scand. **51**, 167 (1997).

[20] J. E. Pollard, D. J. Trevor, J. E. Reutt, and D. A. Shirley, J. Chem. Phys. **81**, 5302 (1984).

[21] K. Somasundram and N. C. Handy, J. Chem. Phys. **84**, 2899 (1986).

[22] J. A. Draves and W. S. Taylor, unpublished.

[23] D. Feller, J. Chem. Phys. **96**, 6104 (1992).

[24] D. Feller, J. Chem. Phys. **98**, 7059 (1993).

[25] T. Helgaker, W. Klopper, H. Koch, and J. Noga, J. Chem. Phys. **106**, 9639 (1997).

[26] A. D. Becke, J. Chem. Phys. **98**, 1372 (1993).

[27] C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).

[28] C. Møller and M. S. Plesset, Phys. Rev. **46**, 618 (1934).

[29] G. D. Purvis and R. J. Bartlett, J. Chem. Phys. **76**, 1910 (1982).

[30] G. E. Scuseria, Chem. Phys. Lett. **176**, 27 (1991).

[31] J. A. Pople, R. Krishnan, H. B. Schlegel, and J. S. Binkley, Int. J. Quantum Chem. Symp. **13**, 225 (1979).

[32] J. Gauss, J. F. Stanton, and R. J. Bartlett, J. Chem. Phys. **95**, 2623 (1991).

[33] J. D. Watts, J. Gauss, and R. J. Bartlett, J. Chem. Phys. **98**, 8718 (1993).

[34] Basis sets were obtained from the Extensible Computational Chemistry Environment Basis Set Database, Version 1/13/03, as developed and distributed by the Molecular Science Computing Facility, Environmental and Molecular Sciences Laboratory with is part of the Pacific Northwest Laboratory , P.O. Box 999, Richland, Washington 99352, USA, and funded by the U.S. Department of Energy.

[35] J. D. Watts and R. J. Bartlett, J. Chem. Phys. **93**, 6104 (1990).

[36] N. C. Handy, J. A. Pople, M. Head-Gordon, K. Raghavachari, and G. W. Trucks, Chem. Phys. Lett. **164**, 185 (1989).

[37] K. Raghavachari, J. A. Pople, E. S. Replogle, and M. Head-Gordon, J. Phys. Chem. **94**, 5579 (1990).

[38] B. J. Johnson and M. J. Fisch, J. Chem. Phys. **100**, 7429 (1994).

[39] J. F. Stanton and J. Gauss, Analytic evaluation of second derivatives of the energy: Computational strategies for the CCSD and CCSD(T) approximations, in *Recent Advances in Coupled-Cluster Methods*, edited by R. J. Bartlett, volume 3 of *Recent Advances in Computational Chemistry*, pages 49–79. World Scientific, Singapore, 1997.

[40] T. D. Crawford, C. D. Sherrill, E. F. Valeev, J. T. Fermann, M. L. Leininger, R. A. King, S. T. Brown, C. L. Janssen, E. T. Seidl, Y. Yamaguchi, W. D. Allen, Y. Xie, G. Vacek, T. P. Hamilton, C. B. Kellogg, R. B. Remington, and H. F. Schaefer III, PSI 3.0, development version, PSITECH, Inc., Watkinsville, GA 30677, U.S.A., 1999.

[41] J. Kong, C. A. White, A. I. Krylov, C. D. Sherrill, R. D. Adamson, T. R. Furlani, M. S. Lee, A. M. Lee, S. R. Gwaltney, T. R. Adams, H. Daschel, W. Zhang, P. P. Korambath, C. Ochsenfeld, A. T. B. Gilbert, G. S. Kedziora, D. R. Maurice, N. Nair, Y. Shao, N. A. Besley, P. E. Maslen, J. P. Dombroski, J. Baker, E. F. C. Byrd, T. V. Voorhis, M. Oumi, S. Hirata, C.-P. Hsu, N. Ishikawa, J. Florian, A. Warshel, B. G. Johnson, P. M. W. Gill, M. Head-Gordon, and J. A. Pople, J. Comp. Chem. **21**, 1532 (2000).

[42] J. F. Stanton, J. Gauss, J. D. Watts, M. Nooijen, N. Oliphant, S. A. Perera, P. G. Szalay, W. J. Lauderdale, S. R. Gwaltney, S. Beck, A. Balková, D. E. Bernholdt, K.-K. B. H. Sekino, P. Rozyczko, C. Huber, and R. J. Bartlett, ACES II, Quantum Theory Project, 1998, Integral packages included are VMOL (J. Almlöf and P. R. Taylor), VPROPS (P. R. Taylor), and a modified version of the ABACUS integral derivative package (T. U. Helgaker, H. J. Aa. Jensen, J. Olsen, P. Jørgensen, and P. R. Taylor).

[43] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gompers, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. W. M. Gill, B. G. Johnson, W. Chen, M. W. Wong, J. L. Andres, M. Head-Gordon, E. S. Replogle and J. A. Pople, Gaussian, Inc., Pittsburgh PA, 1998.

[44] G. Herzberg, *Electronic Spectra and Electronic Structure of Polyatomic Molecules*, volume 3 of *Molecular Spectra and Molecular Structure*. Krieger, Malabar, Florida, reprint edition, 1991.

[45] A. L. L. East and W. D. Allen, J. Chem. Phys. **99**, 4638 (1993).

[46] J. R. Thomas, B. J. DeLeeuw, G. Vacek, and H. F. Schaefer, J. Chem. Phys. **98**, 1336 (1993).

[47] J. R. Thomas, B. J. DeLeeuw, G. Vacek, T. D. Crawford, Y. Yamaguchi, and H. F. Schaefer, J. Chem. Phys. **99**, 403 (1993).

[48] E. F. Valeev, W. D. Allen, H. F. Schaefer, A. G. Császár, and A. L. L. East, J. Phys. Chem. A **105**, 2716 (2001).

[49] Mathematica, version 4.0; Wolfram Research, Inc.: Champaign, IL, 1999.

[50] T. H. Dunning, J. Chem. Phys. **90**, 1007 (1989).

[51] V. Termath, W. Klopper, and W. Kutzelnigg, J. Chem. Phys. **94**, 2002 (1991).

[52] R. A. Kendall, T. H. Dunning, and R. J. Harrison, J. Chem. Phys. **96**, 6796 (1992).

[53] D. E. Woon and T. H. Dunning, J. Chem. Phys. **99**, 1914 (1993).

[54] K. A. Peterson, R. A. Kendall, and T. H. Dunning, J. Chem. Phys. **99**, 1930 (1993).

[55] J. W. Ochterski, G. A. Petersson, and J. A. Montgomery, J. Chem. Phys. **104**, 2598 (1996).

[56] A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen, and A. K. Wilson, Chem. Phys. Lett. **286**, 243 (1998).

[57] J. Olsen, B. O. Roos, P. Jørgensen, and H. J. Aa. Jensen, J. Chem. Phys. **89**, 2185 (1988).

[58] R. J. Harrison and S. Zarrabian, Chem. Phys. Lett. **158**, 393 (1989).

[59] A. V. Luzanov, A. L. Wulfov, and V. O. Krouglov, Chem. Phys. Lett. **197**, 614 (1992).

[60] G. L. Bendazzoli and S. Evangelisti, Int. J. Quantum Chem. Symp. **27**, 287 (1993).

[61] A. O. Mitrushenkov and Y. Y. Dmitriev, Chem. Phys. Lett. **235**, 410 (1995).

[62] A. Povill and J. Rubio, Theor. Chim. Acta **92**, 305 (1995).

[63] S. Evangelisti, G. L. Bendazzoli, R. Ansaloni, and E. Rossi, Chem. Phys. Lett. **233**, 353 (1995).

[64] H. Dachsel, H. Lischka, R. Shepard, J. Nieplocha, and R. J. Harrison, J. Comput. Chem. **18**, 430 (1997).

[65] C. W. Bauschlicher and P. R. Taylor, J. Chem. Phys. **86**, 858 (1987).

[66] C. W. Bauschlicher, J. Phys. Chem. **92**, 3020 (1988).

[67] C. W. Bauschlicher, S. R. Langhoff, H. Partridge, and D. P. Chong, J. Chem. Phys. **89**, 2985 (1988).

[68] J. Olsen, A. M. Sánchez de Meras, H. J. Aa. Jensen, and P. Jørgensen, Chem. Phys. Lett. **154**, 380 (1989).

[69] H. Koch and R. J. Harrison, J. Chem. Phys. **95**, 7479 (1991).

[70] J. M. Galbraith, G. Vacek, and H. F. Schaefer, J. Mol. Structure **300**, 281 (1993).

[71] J. M. Anglada and J. M. Bofill, Theor. Chim. Acta **92**, 369 (1995).

[72] C. D. Sherrill, M. L. Leininger, T. J. Van Huis, and H. F. Schaefer, J. Chem. Phys. **108**, 1040 (1998).

[73] M. L. Leininger, C. D. Sherrill, W. D. Allen, and H. F. Schaefer, J. Chem. Phys. **108**, 6717 (1998).

[74] T. van Mourik and J. H. van Lenthe, J. Chem. Phys. **102**, 7479 (1995).

[75] J. Olsen, P. Jørgensen, H. Koch, A. Balková, and R. J. Bartlett, J. Chem. Phys. **104**, 8007 (1996).

[76] H. Larsen, J. Olsen, P. Jørgensen, and O. Christiansen, J. Chem. Phys. **113**, 6677 (2000).

[77] A. Dutta and C. D. Sherrill, J. Chem. Phys. **118**, 1610 (2003).

[78] T. H. Dunning, J. Chem. Phys. **53**, 2823 (1970).

[79] P. C. Hariharan and J. A. Pople, Theor. Chim. Acta **28**, 213 (1973).

[80] P.-O. Widmark, P.-Å. Malmqvist, and B. O. Roos, Theor. Chim. Acta **77**, 291 (1990).

[81] P.-O. Löwdin, Rev. Mod. Phys. **97**, 1474 (1955).

[82] C. D. Sherrill and H. F. Schaefer, The configuration interaction method: Advances in highly correlated approaches, in *Advances in Quantum Chemistry*, edited by P.-O. Löwdin, volume 34, pages 143–269. Academic Press, New York, 1999.

[83] S. Huzinaga, J. Chem. Phys. **42**, 1293 (1965).

[84] K. P. Huber and G. Herzberg, *Constants of Diatomic Molecules*. Van Nostrand Reinhold, New York, 1979.

[85] A. Carrington and D. A. Ramsey, Phys. Scripta **25**, 272 (1982).

[86] J. M. L. Martin, J. Chem. Phys. **100**, 8186 (1994).

[87] T. Helgaker, J. Gauss, P. Jørgensen, and J. Olsen, J. Chem. Phys. **106**, 6430 (1997).

[88] M. O. Sinnokrot and C. D. Sherrill, J. Chem. Phys. **115**, 2439 (2001).

[89] J. Almlöf, B. J. DeLeeuw, P. R. Taylor, C. W. Bauschlicher, and P. Siegbahn, Int. J. Quantum Chem. Symp. **23**, 345 (1989).

[90] H. F. Schaefer, *Configuration Interaction Wave Functions and the Properties of Atoms and Diatomic Molecules*, PhD thesis, Stanford University, Stanford, CA, 1969.

[91] B. Huron, J. P. Malrieu, and P. Rancurel, J. Chem. Phys. **58**, 5745 (1973).

[92] R. J. Buenker and S. D. Peyerimhoff, Theor. Chim. Acta **35**, 33 (1974).

[93] K. Andersson and B. O. Roos, Multiconfigurational second-order perturbation theory, in *Modern Electronic Structure Theory*, edited by D. R. Yarkony, volume 2 of *Advanced Series in Physical Chemistry*, pages 55–109. World Scientific, Singapore, 1995.

[94] M. R. Hoffmann, Quasidegenerate perturbation theory using effective hamiltonians, in *Modern Electronic Structure Theory*, edited by D. R. Yarkony, volume 2 of *Advanced Series in Physical Chemistry*, pages 1166–1190. World Scientific, Singapore, 1995.

[95] P. G. Szalay, Towards state-specific formulation of multireference coupled-cluster theory: Coupled electron pair approximations (cepa) leading to multireference configuration interaction (mr-ci) type equations, in *Recent Advances in Coupled-Cluster Methods*, edited by R. J. Bartlett, volume 3 of *Recent Advances in Computational Chemistry*, pages 81–124. World Scientific, Singapore, 1997.

[96] U. S. Mahapatra, B. Datta, and D. Mukherjee, A state-specific multi-reference coupled cluster approach for treating quasi-degeneracy, in *Recent Advances in Coupled-Cluster Methods*, edited by R. J. Bartlett, volume 3 of *Recent Advances in Computational Chemistry*, pages 155–181. World Scientific, Singapore, 1997.

[97] L. Adamowicz and J.-P. Malrieu, Multi-reference self-consistent size-extensive configuration interaction (ci) — a bridge bteween the coupled-cluster method and the ci method, in *Recent Advances in Coupled-Cluster Methods*, edited by R. J. Bartlett, volume 3 of *Recent Advances in Computational Chemistry*, pages 307–330. World Scientific, Singapore, 1997.

[98] X. Li and J. Paldus, Mol. Phys. **98**, 1185 (2000).

[99] H. Lischka, R. Shepard, R. M. Pitzer, I. Shavitt, M. Dallos, T. Müller, P. G. Szalay, M. Seth, G. S. Kedziora, S. Yabushita, and Z. Zhang, Phys. Chem. Chem. Phys. **3**, 664 (2001).

[100] K. Andersson and B. O. Roos, Int. J. Quantum Chem. **45**, 591 (1993).

[101] C. W. Bauschlicher, S. R. Langhoff, P. R. Taylor, N. C. Handy, and P. J. Knowles, J. Chem. Phys. **85**, 1469 (1986).

[102] C. W. Bauschlicher and P. R. Taylor, J. Chem. Phys. **86**, 5600 (1987).

[103] C. W. Bauschlicher and P. R. Taylor, J. Chem. Phys. **86**, 1420 (1987).

[104] C. W. Bauschlicher and S. R. Langhoff, Chem. Phys. Lett. **135**, 67 (1987).

[105] C. W. Bauschlicher and P. R. Taylor, J. Chem. Phys. **85**, 2779 (1986).

[106] A. I. Krylov, C. D. Sherrill, E. F. C. Byrd, and M. Head-Gordon, J. Chem. Phys. **109**, 10669 (1998).

[107] A. I. Krylov, C. D. Sherrill, and M. Head-Gordon, J. Chem. Phys. **113**, 6509 (2000).

[108] H. Larsen, J. Olsen, P. Jørgensen, and O. Christiansen, J. Chem. Phys. **114**, 10985 (2001).

[109] W. D. Laidig and R. J. Bartlett, Chem. Phys. Lett. **104**, 424 (1984).

[110] S. R. Gwaltney, C. D. Sherrill, M. Head-Gordon, and A. I. Krylov, J. Chem. Phys. **113**, 3548 (2000).

[111] A. I. Krylov, J. Chem. Phys. **113**, 6052 (2000).

[112] K. Kowalski and P. Piecuch, J. Chem. Phys. **113**, 18 (2000).

[113] K. Kowalski and P. Piecuch, Chem. Phys. Lett. **344**, 165 (2001).

[114] A. I. Krylov and C. D. Sherrill, J. Chem. Phys. **116**, 3194 (2002).

[115] B. O. Roos, P. R. Taylor, and P. E. M. Siegbahn, Chem. Phys. **48**, 157 (1980).

[116] K. Ruedenberg, M. W. Schmidt, M. M. Gilbert, and S. T. Elbert, Chem. Phys. **71**, 41 (1982).

[117] P. Saxe, D. J. Fox, H. F. Schaefer, and N. C. Handy, J. Chem. Phys. **77**, 5584 (1982).

[118] R. S. Grev and H. F. Schaefer, J. Chem. Phys. **96**, 6850 (1992).

[119] D. Z. Goodson and M. Zheng, Chem. Phys. Lett. **365**, 396 (2002).

[120] F. B. Brown and D. G. Truhlar, Chem. Phys. Lett. **113**, 441 (1985).

[121] T. V. Voorhis and M. Head-Gordon, J. Chem. Phys. **112**, 5633 (2000).

[122] G. J. O. Beran, S. R. Gwaltney, and M. Head-Gordon, J. Chem. Phys. **117**, 3040 (2002).

[123] B. O. Roos, Int. J. Quantum Chem. **14**, 175 (1980).

[124] MOLCAS Version 5.2. K. Andersson and M. Barysz and A. Bernhardsson, M. R. A. Blomberg and D. L. Cooper and T. Fleig and M. P. Fülsher and C. de Graaf and B. A. Hess and G. Karlström and R. Lindh and P.-Å. Malmqvist and P. Neogrády and J. Olsen and B. O. Roos and A. J. Sadlej and M. Schütz and B. Schimmelpfennig and L. Seijo and L. Serrano-Andrés and P. E. M. Siegbahn and Jonna Stålring and T. Thorsteinsson and V. Veryazov and P.-O. Widmark, Lund Unversity, Sweden (2000).

[125] T. D. Crawford, C. D. Sherrill, E. F. Valeev, J. T. Fermann, R. A. King, M. L. Leininger, S. T. Brown, C. L. Janssen, E. T. Seidl, J. P. Kenny, and W. D. Allen, PSI 3.2, 2003.

[126] D. L. Gray and A. G. Robiette, Mol. Phys. **37**, 1901 (1979).

[127] K. Andersson, P.-Å. Malmqvist, B. O. Roos, A. J. Sadlej, and K. Wolinski, J. Phys. Chem. **94**, 5483 (1990).

[128] H. J. J. van Dam, J. H. van Lenthe, and P. J. A. Ruttink, Int. J. Quantum Chem. **72**, 549 (1999).

[129] K. Andersson, P.-Å. Malmqvist, and B. O. Roos, J. Chem. Phys. **96**, 1218 (1992).

[130] J. Olsen and M. P. Fülscher, Chem. Phys. Lett. **326**, 225 (2000).

[131] A. Van Orden and R. J. Saykally, Chem. Rev. **98**, 2313 (1998).

[132] T. Juhász and D. A. Mazziotti, J. Chem. Phys. **121**, 1201 (2004).

[133] K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, Chem. Phys. Lett. **157**, 479 (1989).

[134] M. L. Abrams and C. D. Sherrill, J. Phys. Chem. A **107**, 5611 (2003).

[135] M. Boggio-Pasqua, A. I. Voronin, P. Halvick, and J.-C. Rayez, J. Mol. Struct. (Theochem) **531**, 159 (2000).

[136] C. W. Bauschlicher and S. R. Langhoff, J. Chem. Phys. **87**, 2919 (1987).

[137] O. Christiansen, H. Koch, P. Jørgensen, and J. Olsen, Chem. Phys. Lett. **256**, 185 (1996).

[138] M. L. Abrams and C. D. Sherrill, J. Chem. Phys. **118**, 1604 (2003).

[139] P. Piecuch, K. Kowalski, I. S. O. Pimienta, and S. A. Kucharski, Method of moments of coupled-cluster equations: A new theoretical framework for designing "black-box" approaches for molecular potential energy surfaces, in *Low-lying potential energy surfaces*, edited by M. R. Hoffmann and K. G. Dyall, ACS Symposium Series 828, pages 31–64. American Chemical Society, Washington, DC, 2002.

[140] P. Piecuch, K. Kowalski, I. S. O. Pimienta, and M. J. McGuire, Int. Rev. Phys. Chem. **21**, 527 (2002).

[141] W. D. Laidig, P. Saxe, and R. J. Bartlett, J. Chem. Phys. **86**, 887 (1987).

[142] K. B. Ghose, P. Piecuch, and L. Adamowicz, J. Chem. Phys. **103**, 9331 (1995).

[143] W. J. Hehre, R. Ditchfield, and J. A. Pople, J. Chem. Phys. **56**, 2257 (1972).

[144] T. D. Crawford, C. D. Sherrill, E. F. Valeev, J. T. Fermann, R. A. King, M. L. Leininger, S. T. Brown, C. L. Janssen, E. T. Seidl, J. P. Kenny, and W. D. Allen, J. Comp. Chem., submitted .

[145] N. C. Handy, Chem. Phys. Lett. **74**, 280 (1980).

[146] P. Piecuch, V. Špirko, A. E. Kondo, and J. Paldus, J. Chem. Phys. **104**, 4699 (1996).

[147] C. D. Sherrill, Bond breaking in quantum chemistry, Annu. Rev. Comp. Chem., Vol. 1, in press.

[148] C. D. Sherrill and H. F. Schaefer, J. Phys. Chem. **100**, 6069 (1996).

[149] M. W. Schmidt and M. S. Gordon, Ann. Rev. Phys. Chem. **49**, 233 (1998).

[150] H.-J. Werner and P. J. Knowles, J. Chem. Phys. **82**, 5053 (1985).

[151] H. J. Aa. Jensen and H. Ågren, Chem. Phys. **104**, 229 (1986).

[152] G. Chaban, M. W. Schmidt, and M. S. Gordon, Theor. Chem. Acc. **97**, 88 (1997).

[153] R. Shepard, Adv. Chem. Phys. **69**, 63 (1987).

[154] E. R. Davidson, Rev. Mod. Phys. **44**, 451 (1972).

[155] P.-O. Löwdin and H. Shull, Rev. Mod. Phys. **101**, 1730 (1956).

[156] H. Jørgen, A. Jensen, P. Jørgensen, H. Ågren, and J. Olsen, J. Chem. Phys. **88**, 3834 (1988).

[157] P. Pulay and T. P. Hamilton, J. Chem. Phys. **88**, 4926 (1988).

[158] J. M. Bofill and P. Pulay, J. Chem. Phys. **90**, 3637 (1989).

[159] M. S. Gordon, M. W. Schmidt, G. M. Chaban, K. R. Glaesemann, W. J. Stevens, and C. Gonzalez, J. Chem. Phys. **110**, 4199 (1999).

[160] M. L. Abrams and C. D. Sherrill, J. Chem. Phys. **118**, 1604 (2003).

[161] H. J. Aa. Jensen, P. Jørgensen, H. Ågren, and J. Olsen, J. Chem. Phys. **88**, 3834 (1988).

[162] P. Pulay and T. P. Hamilton, J. Chem. Phys. **88**, 4926 (1988).

[163] L. Bytautas, J. Ivanic, and K. Ruedenberg, J. Chem. Phys. **119**, 8217 (2003).

[164] D. M. Potts, C. M. Taylor, R. K. Chadhuri, and K. F. Freed, J. Chem. Phys. **114**, 2592 (2001).

[165] Y.-K. Choe, Y. Nakao, and K. Hirao, J. Chem. Phys. **115**, 621 (2001).

[166] R. Shepard, The analytic gradient method for configuration interaction wave functions, in *Modern Electronic Structure Theory*, edited by D. R. Yarkony, volume 2 of *Advanced Series in Physical Chemistry*, pages 345–458. World Scientific, Singapore, 1995.

[167] J. M. Bofill and P. Pulay, J. Chem. Phys. **90**, 3637 (1989).

[168] R. J. Bartlett, Int. J. Mol. Sci. **3**, 579 (2002).

[169] M. L. Abrams and C. D. Sherrill, J. Phys. Chem. A **107**, 5611 (2003).

[170] P. Pulay, personal communication.

[171] L. M. Cheung, K. R. Sundberg, and K. Ruedenberg, J. Am. Chem. Soc. **100**, 8024 (1978).

[172] N. C. Handy, Chem. Phys. Lett. **74**, 280 (1980).

[173] P. J. Knowles and N. C. Handy, Chem. Phys. Lett. **111**, 315 (1984).

[174] S. Hirata and R. J. Bartlett, Chem. Phys. Lett. **321**, 216 (2000).

[175] J. Olsen, J. Chem. Phys. **113**, 7140 (2000).

[176] M. Kállay and P. R. Surján, J. Chem. Phys. **113**, 1359 (2000).

[177] S. Hirata, J. Phys. Chem. A **107**, 9887 (2003).

[178] M. Kállay and P. R. Surján, J. Chem. Phys. **115**, 2945 (2001).

[179] M. Kállay, P. G. Szalay, and P. R. Surján, J. Chem. Phys. **117**, 980 (2002).

[180] C. D. Sherrill and H. F. Schaefer, J. Phys. Chem. **100**, 6069 (1996).

[181] M. L. Leininger, W. D. Allen, H. F. Schaefer, and C. D. Sherrill, J. Chem. Phys. **112**, 9213 (2000).

[182] C. D. Sherrill and H. F. Schaefer, The configuration interaction method: Advances in highly correlated approaches, in *Advances in Quantum Chemistry*, edited by P.-O. Löwdin, volume 34, pages 143–269. Academic Press, New York, 1999.

[183] B. Roos, Chem. Phys. Lett. **15**, 153 (1972).

[184] M. L. Leininger, C. D. Sherrill, W. D. Allen, and H. F. Schaefer, J. Comput. Chem. **22**, 1574 (2001).

[185] R. J. Bartlett, Coupled-cluster theory: An overview of recent developments, in *Modern Electronic Structure Theory*, edited by D. R. Yarkony, volume 2 of *Advanced Series in Physical Chemistry*, pages 1047–1131. World Scientific, Singapore, 1995.

[186] J. Paldus and X. Li, A critical assessment of coupled cluster method in quantum chemistry, in *Advances in Chemical Physics*, edited by I. Prigogine and S. A. Rice, volume 110, pages 1–175. Wiley, New York, 1999.

[187] T. D. Crawford and H. F. Schaefer, An introduction to coupled cluster theory for computational chemists, in *Reviews in Computational Chemistry*, edited by K. B. Lipkowitz and D. B. Boyd, volume 14, pages 33–136. VCH Publishers, New York, 2000.

[188] I. Shavitt, The method of configuration interaction, in *Methods of Electronic Structure Theory*, edited by H. F. Schaefer, pages 189–275. Plenum Press, New York, 1977.

[189] N. Oliphant and L. Adamowicz, J. Chem. Phys. **94**, 1229 (1991).

[190] N. Oliphant and L. Adamowicz, J. Chem. Phys. **96**, 3739 (1992).

[191] J. Olsen, B. O. Roos, P. Jørgensen, and H. J. Aa. Jensen, J. Chem. Phys. **89**, 2185 (1988).

[192] T. A. Ruden, T. Helgaker, P. Jørgensen, and J. Olsen, Chem. Phys. Lett. **371**, 62 (2003).

[193] M. L. Abrams, DBQC v.1.0, 2005.

[194] J. Olsen, P. Jørgensen, H. Koch, A. Balková, and R. J. Bartlett, J. Chem. Phys. **104**, 8007 (1996).

[195] J. Paldus, M. Takahashi, and B. W. H. Cho, Int. J. Quantum Chem. **18**, 237 (1984).

[196] J. W. Krogh and J. Olsen, Chem. Phys. Lett. **344**, 578 (2003).

[197] G. K.-L. Chan, M. Kállay, and J. Gauss, J. Chem. Phys. **121**, 6110 (2004).

[198] J. Ivanic, J. Chem. Phys. **119**, 9364 (2003).

[199] Y. G. Khait, J. Song, and M. R. Hoffmann, Int. J. Quantum Chem. **99**, 210 (2004).

[200] B. O. Roos, P. Linse, P. E. M. Siegbahn, and M. R. A. Blomberg, Chem. Phys. **1-2**, 197 (1982).

[201] P. Celani and H.-J. Werner, J. Chem. Phys. **112**, 5546 (2000).

[202] H. Nakano, R. Uchiyama, and K. Hirao, J. Comput. Chem. **23**, 1166 (2002).

[203] P. Piecuch, N. Oliphant, and L. Adamowicz, J. Chem. Phys. **99**, 1875 (1993).

[204] P. Piecuch, S. Kucharski, and R. J. Bartlett, J. Chem. Phys. **110**, 6103 (1999).

[205] P. J. Knowles, Chem. Phys. Lett. **155**, 513 (1989).

[206] R. J. Harrison, J. Chem. Phys. **94**, 5021 (1991).

[207] J. C. Greer, J. Chem. Phys. **103**, 1821 (1995).

[208] J. C. Greer, J. Chem. Phys. **103**, 7996 (1995).

[209] M. L. Abrams and C. D. Sherrill, General-order single- and multi-reference configuration interaction and coupled-cluster theory: symmetric dissociation of water, in press.

[210] J. Ivanic and K. Ruedenberg, Theoret. Chim. Acta **106**, 339 (2001).

[211] M. Head-Gordon, T. V. Voohris, S. R. Gwaltney, and E. F. C. Byrd, Coupled cluster methods for bond-breaking, in *Low-lying potential energy surfaces*, edited by M. R. Hoffmann and K. G. Dyall, ACS Symposium Series 828, pages 93–108. American Chemical Society, Washington, DC, 2002.

[212] X. Li and J. Paldus, Simultaneous account of dynamic and nondynamic correlations based on complementarity of CI and CC approaches, in *Low-lying potential energy surfaces*, edited by M. R. Hoffmann and K. G. Dyall, ACS Symposium Series 828, pages 10–30. American Chemical Society, Washington, DC, 2002.

[213] O. Hino, T. Kinoshita, and R. J. Bartlett, J. Chem. Phys. **121**, 1206 (2004).

[214] T. Kinoshita, O. Hino, and R. J. Bartlett, J. Chem. Phys. **119**, 7756 (2003).

[215] R. J. Bartlett, J. Phys. Chem. **93**, 1697 (1989).

[216] C. Hampel and H.-J. Werner, J. Chem. Phys. **16**, 6286 (1996).

[217] P. E. Maslen and M. Head-Gordon, Chem. Phys. Lett. **283**, 102 (1998).

[218] G. E. Scuseria and P. Y. Ayala, J. Chem. Phys. **111**, 8330 (1999).

[219] M. Schütz, Phys. Chem. Chem. Phys. **4**, 3941 (2002).

[220] H. Koch and P. Jorgensen, J. Chem. Phys. **93**, 3333 (1990).

[221] J. F. Stanton and R. J. Bartlett, J. Chem. Phys. **98**, 7029 (1993).

[222] J. F. Stanton, J. Chem. Phys. **101**, 8928 (1994).

[223] J. S. Arponen, R. F. Bishop, and E. Pajanne, Phys. Rev. A **36**, 2519 (1987).

[224] J. S. Arponen, R. F. Bishop, and E. Pajanne, Phys. Rev. A **36**, 2539 (1987).

[225] W. A. Lester, editor, *Recent Advances in Quantum Monte Carlo Methods*, volume 2 of *Recent Advances in Computational Chemistry*. World Scientific, Singapore, 1997.

[226] S. Lu, J. Chem. Phys. **118**, 9528 (2003).

[227] A. Aspuru-Guzik, O. E. Akramine, J. C. Grossman, and W. A. Lester, J. Chem. Phys. **120**, 3049 (2004).

[228] S. Lu, J. Chem. Phys. **121**, 10365 (2004).

[229] G. K.-L. Chan and M. Head-Gordon, J. Chem. Phys. **116**, 4462 (2002).

[230] G. K.-L. Chan and M. Head-Gordon, J. Chem. Phys. **118**, 8551 (2003).

[231] G. K.-L. Chan, J. Chem. Phys. **120**, 3172 (2004).

[232] D. A. Mazziotti, Phys. Rev. A **57**, 4219 (1998).

[233] D. A. Mazziotti, Phys. Rev. A **60**, 4396 (1999).

[234] D. A. Mazziotti, J. Chem. Phys. **115**, 8305 (2001).

[235] D. A. Mazziotti, Phys. Rev. A **65**, 062511 (2002).